The Dissertation Committee for Yaqi Wan Certifies that this is the approved version of the following dissertation:

Probing Stability, Specificity, and Modular Structure in Group I Intron RNAs

Committee:

Rick Russell, Supervisor

Karen Browning

Robin Gutell

David Hoffman

Pengyu Ren

Probing Stability, Specificity, and Modular Structure in Group I Intron RNAs

by

Yaqi Wan, B. S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2010

Dedication

To my supporting parents and friends

Acknowledgements

I would like to take this opportunity to show my gratitude to those who made this thesis possible. I would like to express my sincerest gratitude to my supervisor, Rick Russell for his excellent guidance, advice and patience throughout my graduate school time. I was inspired so much from his scientific intuition and logical thinking.

I gratefully acknowledge Dr. Robin Gutell for providing the preliminary data for one of my projects and great help on my research for the past several years. I would also like to thank Dr. David Hoffman for answering my questions and helping me develop a background in biophysics; Dr. Karen Browning and Dr. Pengyu Ren for valuable advice during my talks.

Many thanks go to all my lab colleagues for creating an excellent environment to conduct research. I would like to thank previous lab members, Hari Bhaskaran, Travis Johnson and Amanda Chadee, who helped me to get started in the Russell lab. I thank Pilar Tijerina for the great help during the data collecting days in the Argonne National Laboratory in Chicago. I would also like to thank all the current lab members, Brian Canon, David Mitchell, Jeffrey Potratz, Cynthia Pan and Inga Jarmoskaite for their valuable suggestions in my group meetings and thesis writing.

Last, but not least, I would like to thank my parents, my boyfriend and my great friends in Austin. They have always supported and encouraged me through the good and bad times in the past.

Probing Stability, Specificity, and Modular Structure in Group I Intron RNAs

Publication No._____

Yaqi Wan, Ph.D.

The University of Texas at Austin, 2010

Supervisor: Rick Russell

Many functional RNAs are required to fold into specific three-dimensional structures. A fundamental property of RNA is that its secondary structure and even some tertiary contacts are highly stable, which gives rise to independent modular RNA motifs and makes RNAs prone to adopting misfolded intermediates. Consequently, in addition to stabilizing the native structure relative to the unfolded species (defined here as stability), RNAs are faced with the challenge of stabilizing the native structure relative to alternative structures (defined as structural specificity). How RNAs have evolved to overcome these challenges is incompletely understood.

Self-splicing group I introns have been used to study RNA structure and folding for decades. Among them, the *Tetrahymena* intron was the first discovered and has been studied extensively. In this work, we found that a version of the intron that was generated by *in vitro* selection for enhanced stability also displayed enhanced specificity against a stable misfolded structure that is globally similar to the native state, despite the absence of selective pressure to increase the energy gap between these structures. Further

dissection suggests that the increased specificity against misfolding arises from two point mutations, which strengthen a local tertiary contact network that apparently cannot form in the misfolded conformation. Our results suggest that the structural rigidity and intricate networks of contacts inherent to structured RNAs can allow them to evolve exquisite structural specificity without explicit negative selection, even against closely-related alternative structures.

To explore further how RNAs gain stability from intricate architectures, we examined a novel group I intron from red algae (*Bangia*). Biochemical methods and computational modeling suggest that this intron possesses general motifs of group IC1 introns but also forms an atypical tertiary contact, which has been reported previously in other subgroups and helps position the reactive helix at the active site. In the *Bangia* intron, the partners have been swapped relative to known group I RNAs that include this contact. This result underscores the modular nature of RNA motifs and provides insight into how structured RNAs can arrange helices and contacts in multiple ways to achieve and stabilize functional structures.

Table of Contents

List of Tables xi

List of Figures xii

Chapter 1: Overview of RNA structure and folding 1

1.1 Introduction 1

1.2 RNA architecture 1

1.2.1 Diversity of structured RNAs 2

1.2.2 Structural features that contribute to the RNA folding problem .. 4

1.3 Folding of structured RNAs 5

1.3.1 RNA misfolding *in vitro* 6

1.3.2 RNA folding and misfolding *in vivo* 8

1.4 A model system for studying RNA structure and folding: group I introns 11

1.4.1 Structural features of the *Tetrahymena* group I ribozyme 12

1.4.2 Folding pathway of the *Tetrahymena* group I ribozyme 12

1.4.3 Lessons from other group I RNAs 15

1.4.3.1 Group I introns share a conserved core 15

1.4.3.2 Stability of group I introns relies on the well-organized peripheral elements 15

1.4.3.3 Misfolding of other group I introns 17

1.4.3.4 Novel group I RNAs extend the knowledge of RNA study 18

Chapter 2: Enhanced specificity against misfolding in a thermostable mutant of the *Tetrahymena* ribozyme 29

2.1 Introduction 29

2.2 Materials and methods 31

2.2.1 Preparation of RNA 31

2.2.2 Catalytic activity assay to monitor native ribozyme formation . 32

2.2.3 Mg^{2+} dependence of native ribozyme formation 33

2.2.4 P5abc binding and dissociation 34

2.2.5 Calculation of native specificity in the full-length ribozyme 35
2.2.6 Hydroxyl radical footprinting using Fe (II)-EDTA 35
2.2.7 Thermal denaturation..... 37
2.3 Results..... 37
2.3.1 Increased native specificity in the thermostable R14C mutant.. 37
2.3.2 The peripheral element P5abc is required for increased stability but not specificity..... 40
2.3.3 Two mutations are responsible for the increased specificity..... 42
2.3.4 Dissection of independent energetic effects in the absence of P5abc 44
2.3.5 Specific native state stabilization by 269G/304G in the full-length ribozyme 46
2.4 Discussion 47

Chapter 3: Multiple unfolding events during native folding of the *Tetrahymena* group I ribozyme 67
3.1 Introduction 67
3.2 Materials and Methods..... 69
3.2.1 Materials 69
3.2.2 Fast Fenton footprinting 69
3.3 Results..... 71
3.4 Discussion 74
3.4.1 Escape from I_{trap} via local unfolding 75
3.4.2 Multiple unfolding steps in continued folding from intermediates to the native structure 76

Chapter 4: Extensive Tertiary Contact Formation and Novel Architecture in a Group I RNA..... 85
4.1 Introduction 85
4.2 Materials and methods 88
4.2.1 Preparation of the wild-type and mutant *Bangia* ribozymes 88
4.2.2 Cleavage activity assays 88
4.2.3 Static small angle X-ray scattering..... 89
4.2.4 3D modeling of the *Bangia* ribozyme 89

4.2.5 Simulating ideal solvent accessibility data and SAXS data 90
4.3 Results 90
4.3.1 The *Bangia* ribozyme possesses most of the characteristic tertiary motifs of group I introns 90
4.3.2 Evolution kept the tertiary interaction but swapped the elements 95
4.3.3 Three-dimensional model of the *Bangia* ribozyme 98
4.4 Discussion 103

Chapter 5: The misfolding of a *Bangia* group I ribozyme follows a rapid collapse 123
5.1 Introduction 123
5.2 Methods 125
5.2.1 Time-resolved small angle X-ray scattering 125
5.2.2. Folding of the *Bangia* ribozyme monitored by hydroxyl radical footprinting 125
5.2.3 Activity assay to follow the formation of native ribozyme 126
5.3 Results 126
5.3.1 Small angle X-ray scattering revealed a trapped misfolded construct 126
5.3.2 The folding of the *Bangia* ribozyme involves multiple intermediates 127
5.3.3 Formation of native ribozyme monitored by catalytic activity assay 128
5.4 Discussion and future plan 130

Appendix 139
A.1 DMS footprinting 139
A.2 Transition from Itrap to N requires less structural disruption than from M 140

List of abbreviations 144

References: 145

Vita 156

List of Tables

Table 2.1: Kinetic and thermodynamic constants of P5abc-deleted ribozymes 65

Table 2.2: Kinetics and thermodynamic constants of full-length mutant ribozymes 66

Table 4.1: Tertiary stabilization from the *Bangia* ribozyme on the substrate 120

Table 4.2: Activity measurements for tertiary contact disruption mutants. 121

Table 4.3: Activity measurements for L8/P2 double mutant and L5/P2 double mutant . 122

Table 5.1: Folding of individual regions in the *Bangia* ribozyme mmonitored by hydroxyl radical footprinting 133

List of Figures

Figure 1.1: Structure of small RNAs 20
Figure 1.2: Tertiary structure of large catalytic RNAs/RNP 21
Figure 1.3: Difference between RNA and protein chains 22
Figure 1.4: Folding landscaped of RNA and protein 23
Figure 1.5: Self-splicing group I intron and its engineered ribozyme 24
Figure 1.6: Structure of the model system: *Tetrahymena* group I ribozyme 25
Figure 1.7: Model of Tetrahymena ribozyme folding pathway and alternative intermediate 26
Figure 1.8: The overlapping of three group I intron crystal structures 27
Figure 1.9: Arrangements of peripheral elements in group I introns to buttress the core 28
Figure 2.1: Thermodynamic cycle used to calculate the equilibrium between the native and misfolded forms 51
Figure 2.2: Secondary structure and tertiary structural model of the R14C ribozyme 52
Figure 2.3: Increased specificity for native folding in the R14C$^{\Delta P5abc}$ mutant ribozyme 53
Figure 2.4: Hydroxyl radical footprinting of native and misfolded conformers of the full length R14C ribozyme 54
Figure 2.5: Preferential binding of P5abc to the native R14C$^{\Delta P5abc}$ ribozyme 55
Figure 2.6: Native stabilization by the R14C mutations requires P5abc 56
Figure 2.7: Folding transitions of the full-length and P5abc-deleted ribozymes 57
Figure 2.8: Refolding of full-length ribozyme variants 58
Figure 2.9: Mg^{2+} dependence of native ribozyme formation 59
Figure 2.10: Thermodynamic cycle for the A269G/A304G ribozyme variant 60

Figure 2.11: Only three substitutions affect stability and specificity in the absence of P5abc 61
Figure 2.12: A reversion mutant G269A/G304A in the R14C background loses the enhanced specificity for native folding 62
Figure 2.13: A reversion mutant G269A/G304A in the R14C background loses the enhanced specificity for native folding 63
Figure 2.14: Physical model for the enhanced specificity from the A269G and A304G mutations 64
Figure 3.1: Secondary structure of the *Tetrahymena* group I ribozyme 78
Figure 3.2: Folding schemes for the *Tetrahymena* ribozyme under different temperatures 79
Figure 3.3: Hydroxyl radical and DMS footprinting of the I_{trap} folding intermediate 81
Figure 3.4: Hydroxyl radical and DMS footprinting comparisons of the native state and unfolded state 82
Figure 3.5: Comparisons of footprinting pattern for the native ribozyme and earlier intermediates displayed on the secondary structure 83
Figure 3.6: Model for continued folding of the intermediates I_{trap} and M 84
Figure 4.1: Secondary structure of the *Bangia* group I intron 107
Figure 4.2: Measurement of catalytic activity of the wild-type *Bangia* ribozyme 108
Figure 4.3: Varying conditions to confirm that the cleavage rate is rate-limiting in the catalytic activity assay 109
Figure 4.4 Mg^{2+} dependent folding of the *Bangia* ribozyme monitored by SAXS 110
Figure 4.5: Effects of tertiary element mutations on catalytic activity 111
Figure 4.6: Comparison of folded *vs.* unfolded *Bangia* ribozyme by hydroxyl radical footprinting 112

Figure 4.7: Mutations in the L8 and L5b.1 loop affect activity differently 113
Figure 4.8: Double mutant cycle of L8UUCG mutant and P2UA mutant 114
Figure 4.9: Design of rescue mutants to confirm the contact between the L8 tetraloop and the P2 receptor 115
Figure 4.10: Threading of the *Bangia* ribozyme on the structures of the *Tetrahymena* and *Azoarcus* group I ribozymes 116
Figure 4.11: Computational models of *Bangia* ribozyme generated by RNABuilder 117
Figure 4.12: Cleavage activity of P2 shortening mutants..... 118
Figure 4.13: Comparison between the ideal solvent accessibility data of 3D models and the hydroxyl footprinting data 119
Figure 5.1: Compaction of the *Bangia* ribozyme monitored by SAXS 134
Figure 5.2: SAXS reveals partially folded intermediates of the *Bangia* ribozyme..... 135
Figure 5.3: Time-dependent protection of residue 174 monitored by hydroxyl radical footprinting 136
Figure 5.4: Formation of the native ribozyme monitored by activity assays 137
Figure 5.5: Urea dependence of refolding kinetics of the *Bangia* ribozyme..... 138
Figure A1: Acceleration by urea of I_{trap} folding to the native state 142
Figure A2: Effects of tertiary contact disruptions on re-folding of I_{trap} and M 143

Chapter 1: Overview of RNA structure and folding

1.1 INTRODUCTION

RNAs have been revealed to execute diverse important biological functions in the cell in addition to carrying genetic information. It has been shown that RNAs are involved in translation, tRNA maturation, chromosomal end maintenance, gene regulation and RNA interference (*1-5*). Indeed, only a small portion of transcribed RNAs encode proteins inside the cell, though the entire genome has been transcribed. The remaining non-coding RNAs may undertake more unknown biological functions (*6*).

Prior to functioning many RNAs must fold into well-defined three-dimensional structures. This introduces a fundamental 'RNA folding problem', which can be divided into two aspects. First, RNA must stabilize its native structure relative to all alternative structures in order to specify a unique functional conformation at equilibrium. Second, RNA must escape from the alternative structures on a time scale short enough to function (*7, 8*). Due to the intrinsic properties of RNA structure, it is found that RNAs are prone to adopting stable misfolded structures, and these structures can be long-lived (*7, 9*). Thus, addressing these challenges is a key early step in the functioning of all structured RNAs.

1.2 RNA ARCHITECTURE

To understanding how RNA folds, it is essential to recognize the architecture of the folded RNA. The current view of RNA structure is that it is built upon secondary structure elements such as A-form helices, internal loops and hairpins, which are

arranged into a global structure by tertiary interactions. This is defined as a hierarchical construction because the secondary structure is highly stable and forms independently of tertiary structure (*9-11*). From as simple as 'L-shape' tRNA to as intricate as ribosomal RNA, structured RNAs present a diversity of shapes and organizations. However, there are common themes shared by them that not only impact RNA structure but also folding.

1.2.1 Diversity of structured RNAs

The first recognized structured RNA was tRNA, which is well known for its cloverleaf secondary structure and L-shaped tertiary structure (*12, 13*). In the secondary structure of tRNA, four stem-loops branch out from a junction and are readily formed in the absence of any tertiary contact (*12*). These stems are arranged in the tertiary structure such that each arm of the 'L' is formed from two of the coaxial stacked helices. The two arms, or helical domains are connected by tertiary interactions between the loops at their ends (*13*) (Figure 1.1A). The organization of coaxially stacked helices is frequently seen in other small structured RNAs as well, but the manner may differ from each other.

Indeed, the hairpin ribozyme contains helices and a four-way junction in the secondary structure, however, its architecture differs from tRNA. The stacked helices form two intertwined helical domains that mimic an X-shape (Figure 1.1B) (*14*). Instead of an end-loop interaction, these two domains are connected by internal loops. Another small ribozyme, the hammerhead ribozyme is composed of three stems and a covalent junction, which together adopt a 'Y-shape' (*15*). The helical junction plays an important role in these small ribozymes presumably because it controls the proper coaxial stacking and the arrangement of helices in space (*10*). Additional evidence for the importance of

junctions came from an emerging class of genetic control elements termed riboswitches, in which the helical junction serves as the metabolite binding site and transforms the structure of the entire molecule upon substrate binding (*5*).

Large structured RNAs, including group I and group II introns, RNase P RNA and RNAs in the ribosome subunits, further provide a wealth of information on RNA structure (Figure 1.2). The group I intron from *Tetrahymena thermophila* was the first RNA found to have catalytic activity and has been studied extensively on its structure and folding. The core elements of this intron are two linked helical domains, which are wrapped by exterior peripheral helices and tertiary contacts formed between them (*16, 17*). Functionally similar to group I introns in excising themselves out of pre-mRNA, the group II RNAs possess six domains that radiate from a central junction loop in the secondary structure. While five of the domains are merely single stem-loops, domain 1 consists of several helices linked by junctions and internal loops (*18, 19*). In the tertiary structure, coaxial stacking again dominates in the group II RNAs, with the intricate domain 1 serving as the scaffold. Another example of large structured RNA is RNase P RNA, which promotes the cleavage of the 5' end of pre-tRNA (*20*). Bacterial RNase P RNAs are active *in vitro* on their own, while those from eukaryotes and most archaea require the presence of proteins (*21*). Generally, RNase P RNA has two domains, a catalytic domain (C) and a specificity domain (S). Analogous to group I introns, each of the two major domains in RNase P RNA contains a conserved core and surrounding peripheral helices (*22*). Large structured RNAs have evolved extensively to pack more helices into a single global structure. Inside the most intricate and enormous machine in the cell, the 4500-nucleotide long ribosomal rRNAs organize various helices elegantly by

coaxial stacking and long-range tertiary contacts such as tetraloop-receptors and ribose zippers (*23, 24*).

1.2.2 Structural features that contribute to the RNA folding problem

Despite the variety of RNA structures, there are fundamental characteristics that underlie RNA architecture and govern its folding. The primary sequence of RNA only contains four different types of bases. They can be further divided into two classes: purine and pyrimidine. Compared to proteins, which have 20 unique side chains, the information content in RNA polymers is extremely limited and makes them difficult to specify a single functional structure (*7*). Additionally, A-form RNA helices internalize the bases through base-pairing interactions, which gives the segments a similar appearance from the outside and complicates the formation of a unique structure. This is in contrast to the protein α-helix or β-sheet, which present all of the side chains outward (*7, 9, 25*) (Figure 1.3). The above characteristics highlight the thermodynamic challenge in RNA folding.

The kinetic problem in folding arises from several aspects of RNA structure. First, the independently formed RNA helix is very stable due to the base pairing interaction and the base stacking interaction between adjacent nucleotide bases. Dissociation of an RNA helix with 10 base pairs requires more than an hour, while the α-helix in a protein takes less than microseconds to disorder (*26*). Therefore, RNA can be trapped in an alternative conformation for a long time. Furthermore, the orientation of the RNA backbone is restricted. Although there are six rotable torsion angles inside each residue that presumably allow free movement of the sugar pucker and base, studies have shown that

this range is highly restricted, and only a small fraction of conformations are favored in natural RNAs (*27, 28*). This local rigidity increases the energy barrier between different conformations and can impede the process of escaping from alternative conformations.

1.3 FOLDING OF STRUCTURED RNAS

Stable and long-lasting misfolded intermediates lead to a rugged free energy landscape for RNA folding. Protein, the other biopolymer, folds by excluding solvent from the hydrophobic core. In order to derive an enthalpy larger than the entropy lost, extensive interactions have to be formed simultaneously in many proteins, while partially folded or misfolded structures are not stable (Figure 1.4). On the other hand, RNA has stable partially folded or misfolded intermediates due to the enthalpy contribution from stable secondary structure. Different from the 'two-state' folding in proteins, RNA folding traverses downhill a rough path to the specific functional structure (Figure 1.4). The native structure not only needs to be more stable than the unfolded species, termed traditionally as the '(thermo) stability'; but it also needs to have a lower free energy level than all of the alternative structures, which is termed 'structural specificity'. The complexity of this landscape increases with the size and shape of individual RNAs.

The study of RNA folding has been extensively performed *in vitro* and provides us fruitful information. Unfolded RNA is expanded due to the electrostatic repulsion generated from negatively charged phosphate groups in each monomer. It can readily form secondary structures in the presence of monovalent metal ions. To fold into tertiary structure, most RNAs need divalent ions, especially Mg^{2+} ions, which can nonspecifically bind and neutralize the negatively charged molecules (or 'ion atmosphere') (*29*). The

energetic contribution of specific binding is far less significant. Mounting data have shown that this large scale charge neutralization occurs during the early stages of folding, and it is presumably represented by a rapid and nonspecific collapse that is independent of any tertiary contact formation (e.g. millisecond-scale in group I RNA) (*30, 31*). The collapsed state is usually not the native structure. Therefore, RNA has to search for the active conformation by local rearrangement or multiple unfolding and refolding processes (see Chapter 3 and ref. (*30, 32*)). The difficulty of obtaining the active conformation is thus a major challenge for RNA folding *in vitro*. The *in vivo* folding of RNA is more mysterious because of the complicated environment within cells (see Section 1.3.2), but the underlying principle is definitely shared by both processes.

1.3.1 RNA misfolding *in vitro*

Alternative conformations have been discovered in almost all structured RNAs *in vitro* (*9, 33*). Soon after the first tRNA secondary structure was resolved, investigators noticed that this RNA adopts a long-lived inactive conformation (*34, 35*). Its activity can be restored by incubating at higher temperature, demonstrating that the lack of activity was not caused by damage to the RNA (*35*). Small ribozymes that share similar size with tRNA also have misfolded species. For example, in the catalytic activity assay of the hairpin ribozyme, both steady state and pre-steady state kinetics revealed an inert species before the formation of the functional species (*36*). Evidence of alternative conformation in small RNAs with simple structures suggests the prevalence of misfolding in RNAs.

The first large RNA shown to misfold was the RNase P RNA. Its misfolded conformation was revealed by a kinetic lag in the cleavage activity assay. This incorrect

form can convert to the correct one by pre-incubation at elevated temperature (*20, 37*). The misfolding of the *Tetrahymena* group I intron and its ribozyme version were revealed by catalytic and gel shift assays (*38-40*), which detect misfolded structure by different strategies. While catalytic assays report a decrease in activity to signify alternative structures, gel shift assays probe the existence of misfolded conformations because non-native structures presumably interfere with the correct packing and result in less compact molecules. Recently established single-molecule FRET and small angle X-ray scattering (SAXS) methods have identified misfolded structures based on the same rationale.

Following the discovery of alternative folding, the details of misfolded structure for various RNAs have been dissected. The alternative structure in tRNA was initially showed to have fewer base pairs and a larger size (*41-43*), and was revealed to have more unpaired nucleotides by using complementary oligonucleotide hybridization (*44*). Later, a footprinting assay suggested mismatches between the D-loop and T-loop in the misfolded tRNA (*45*). Although these experiments used different tRNAs, all of them proposed non-native secondary structures as the origin of misfolding. Analogously, an extensive misfolded structure was demonstrated to originate from the coaxial stacking of two helices in a minimal version of the hairpin ribozyme, while these two stems bent at the junction and dock to each other in the native structure (*36, 46*). Indeed, misfolded secondary structure dominates the small RNA folding, which is in accordance with the small number of tertiary interactions in those RNAs (*47, 48*).

This has been confirmed by the complicated misfolded structures in large-size RNAs. The *Tetrahymena* group I intron can be misfolded due to the alternative base pairing between the internal guide sequence and an upstream compete sequence in 5'exon (*49*),

while its ribozyme version folds through more than one misfolded intermediates, which was revealed by complementary oligonucleotide probing in the Williamson lab (details discussed in Section 1.4.2) (*50, 51*). Misfolding of RNase P RNP was studied by the Williamson group via oligonucleotide hybridization as well. The formation of a junction helix between the catalytic domain and specificity domain was shown to be rate limiting (*52*). RNase P RNA folding has been delicately characterized by the Pan and Sosnick groups. They first utilized circular dichroism and UV absorbance spectroscopy to follow RNA folding and suggested that a bacterial RNase P RNA was trapped in kinetic intermediates because the folding can be accelerated by denaturant (*53*). Details of the kinetic trap were disclosed by a circularly permuted ribozyme that contains a break between the catalytic and specificity domain. This mutant folds nearly 15-fold faster than the wild-type ribozyme, suggesting that the origin of misfolding is close to the junction region (*54*). This agrees with the previous oligonucleotide hybridization data (*52*). Later work from this group focused on the independent folding of the catalytic domain. They found that this domain folds more rapidly than the whole ribozyme and is not affected by denaturant, supporting the role of the interdomain junction in misfolding of the full-length ribozyme (*55, 56*).

1.3.2 RNA folding and misfolding *in vivo*

How RNA folds *in vivo* has to be ultimately related with its biological function. Most cellular RNAs have to accomplish their functions in a time scale as short as seconds, which is in contradictory to the prevalence and persistence of misfolding observed *in vitro*. There must be some components in cells that assist RNA folding by either avoiding

misfolded structures or escaping rapidly from kinetic traps. Evidence for different folding *in vitro* and *in vivo* came from the *Tetrahymena* group I intron. The intron accomplishes folding and catalysis in a few seconds inside the organism, the same time scale as the synthesis of this RNA (*57*), while its folding *in vitro* takes much longer (*58, 59*).

RNA folds co-transcriptionally *in vivo* and therefore is influenced by transcription speed and site-specific pausing of the RNA polymerase. In addition, a variety of RNA binding proteins and chaperones may interact with nascent RNA fragment or the complete chain and alter its folding (*60*). To specifically examine the power of co-transcriptional folding and RNA polymerase pausing, the Pan and Sosnick groups have performed complicated studies on co-transcriptional folding of RNase P RNA *in vitro* (*60-62*). The RNA was circularly permuted so that the catalytic domain would be transcribed first, followed by the specificity domain. It was found that the catalytic domain forms an interdomain mispairing with the 5' region of the specificity domain, which impedes folding. In contrast, Mg^{2+}-induced *in vitro* folding generates two independently misfolded domains. Moreover, they found that upon addition of an *Escherichia. Coli* elongation factor NusA, the RNA polymerase stayed longer at a specific pausing site in the catalytic domain, and this avoided the misfolding with the specificity domain (*61*). Together, the results suggest that RNA folds sequentially and that transcription and RNA polymerase pausing could completely change the folding pathway.

The idea that RNA-binding proteins may help RNAs overcome the folding problem *in vivo* was proposed because of the discovery of protein-assisted RNA folding *in vitro* (*7, 63, 64*). A number of RNA chaperones have been identified, such as the ubiquitous

DExD/H-box proteins that can nonspecifically unwind short RNA helices (*9, 65*). The actual function for many of them remains to be determined *in vivo*.

Recent studies carried out by Fedor and her colleagues on the hairpin ribozyme have provided some details about *in vivo* folding. They used a simplified hairpin ribozyme that only contains two stem loops, and introduced alternative structures by complementary insertions upstream or downstream from the ribozyme. Both misfolded constructs were designed to be more stable than the native ribozyme. Intriguingly, both constructs misfolded during *in vivo* folding, suggesting that either the folding is non-sequential *in vivo* or other intracellular factors like RNA chaperones may accelerate the thermodynamic equilibration (*66*). The former possibility was ruled out because the ribozyme activity was restored in a mutant that contains a more stable native duplex, despite that the potential downstream misfolded duplex has greater thermostability (*67*). The results indicated that folding is also sequential *in vivo*; and RNA-binding proteins might decrease the energy barrier between different conformations. But, when the energy barrier is tremendously high, RNA would still be trapped in the early-formed conformation for a long time. The Fedor group also tried to address whether protein chaperones help folding by generally destabilizing the kinetic trap or stabilizing the transition state inside cells. They evaluated the stability of the kinetic trap for this hairpin ribozyme by the duplex dissociation rate constant and found it was the same *in vivo* and *in vitro* (*67, 68*). Thus, they proposed that RNA chaperones facilitate rapid exchange between the conformations by lowering the free energy level of the transition state, presumably via branch migration inside the hairpin ribozyme (*67*). Additional work is required to further determine the role of RNA chaperones in the cell.

1.4 A MODEL SYSTEM FOR STUDYING RNA STRUCTURE AND FOLDING: GROUP I INTRONS

Group I introns are structured elements that can self-splice from the precursor RNA. They have been widespread in many organisms such as fungi, algae and eukaryotes (*69*). The discovery of the catalytic group I intron in *Tetrahymena thermophila* evoked the interest in RNA structure and function (*70*). A growing number of group I introns have been identified from different organisms. There are more than 2000 group I introns, which can be categorized into at least 5 groups and 13 subgroups (*71-73*). They catalyze a two-step transterification reaction with the assistance of an exogenous cofactor guanosine (*40*) (Figure 1.5A). In the first step, the 3'-hydroxyl of the guanosine binds and attacks at the 5' splice site. This transterification reaction results in a 3'-hydroxyl group in the exon and attaches the guanosine to the 5' end of the intron. In the second step, the 3'-hydroxl of the 5'-exon end attacks the 3'-splice site, releasing the ligated exons from the intron. An engineering step was performed later to remove the flanking exons from the intron, the remaining RNA mimics the first step of splicing and cleaves the substrate in *trans* like a protein enzyme *(74)* (Figure 1.5B). To function, group I RNAs must fold into a specific three-dimensional structure, which allows them to become a good model to study RNA structure and folding.

From comparative analysis and various biochemistry experiments, it is now known that the secondary structure of all group I RNAs are constituted of several helices linked by single-stranded junctions or internal loops. They are classified by the sequence of the conserved helices (P4-P6 and P3-P7) and the different organization of the peripheral helices, which include P2, P2.1, P5abc, P9, P9.1, P9.2 and P7.1/P7.2. A few group I

ribozymes have been studied extensively in terms of their architectures and folding, while the *Tetrahymena* group IC1 ribozyme is the most studied prototype.

1.4.1 Structural features of the *Tetrahymena* group I ribozyme

The three-dimensional structure of the *Tetrahymena* ribozyme was revealed by crystal structures of major domains in this ribozyme and a structural model (*75-78*) (Figure 1.6). Two coaxially stacked helices P4-P6 and P3-P7 interact with each other and constitute the conserved core. The interaction surface between them serves as the active center, where the P1 duplex docks and binds the exogenous guanosine. Peripheral helices (P2/P2.1, P5abc, P9, and P9.1/9.2) pack around the core to buttress it by forming long range tertiary contacts. In a more detailed view, P2.1 and P2 coaxially stack on each other, while P2 contacts P5c via base pair interaction on one end (P14) and P2.1 contacts P9.1 on the other end (P13) of the RNA. This long P2/P2.1 duplex, together with the stacked P9.1/P9.2 and the P5abc domain, form a closed circle outside the core (*16, 75, 76*). Other tertiary contacts are formed between L5b and J6/6a, P5bulge and J4/5 as well as L9 and P5.

1.4.2 Folding pathway of the *Tetrahymena* group I ribozyme

The study of *Tetrahymena* ribozyme folding began in 1994, Williamson and his co-workers followed the available unpaired regions during folding by using several DNA oligos that complementary bind to different parts of the RNA. They found that P4-P6 forms before the packing of P3-P7 core, which becomes the rate-limiting step (0.7 min^{-1}). More importantly, the data revealed the existence of intermediates because the rate-

limiting folding rate was independent of Mg^{2+} concentration (*50*). Details of the folding order were further elucidated by Sclavi et al. in 1998 (*79*). They used hydroxyl radicals to probe the solvent accessibility change along with RNA folding. The experimental setup allowed observation on the milliseconds scale, so that fast folding events that could not be visualized by standard bench work were shown. It was proposed that the *Tetrahymena* ribozyme folds in a hierarchical pathway from P5abc, P4-P6, P2/P2.1/P9.1 to P3-P7 with the rate ranging from $2s^{-1}$ to $0.02s^{-1}$. The latter is coincident with the P3-P7 folding observed previously (*50*).

At the same time, the Woodson group used chemical modification and footprinting, combined with mutagenesis to reveal the structure of intermediates during folding. They detected a long-lived misfolded conformation and proposed that a non-native secondary structure instead of P3 was formed within the core element in this misfolded intermediate and termed it alt P3 (Figure 1.7A) (*38*). Mutagenesis and catalytic activity assays performed by different groups found that disrupting the native tertiary contacts in P4-P6 or P13 and P14 accelerated the re-folding of this intermediate towards the final native conformation, whereas the denaturant urea also facilitated overall folding (*38, 51, 59, 80-82*). These data indicated that the long-lived misfolded structure is protected by native interactions, which need to be disrupted before the transition from the misfolded structure to the native one.

The Herschlag and Russell groups designed a quantitative catalytic activity assay to study the complex folding pathway and structured intermediates of the *Tetrahymena* ribozyme (*58, 83, 84*). They demonstrated that a partition (or commitment) point between folding to the native state or the long-lived misfolded states occurs late in the folding

pathway (*83*). Together with the previous data (*59, 81, 82, 85*), they established a folding pathway that the unfolded ribozyme collapses into a compact and structured intermediate I_{trap} rapidly (60 min^{-1}) (rates measured under 37 °C), which then folds with a slower rate (1.5 min^{-1}) into a intermediate $I_{commitment}$. Ten percent of this intermediate directly folds into the native state (≥15 min^{-1}), while the remaining fraction partitions into a stable misfolded species (*83*). This inactive form slowly transfers into the native form with a rate of 0.0014 min^{-1} at 37°C. Therefore, it seems that the previous oligonucleotide hybridization assay illuminated the early steps in the pathway due to experimental time scale (*50*). Indeed, the Williamson group found later that the overall folding followed by cleavage activity assay was much slower than 0.7 min^{-1} (*51*).

The slow transition from the misfolded structure (M) to the native state (N) at room temperature allowed researchers to populate it and investigate its properties further (*83*). Hydroxyl radical footprinting and small angle X-ray scattering (SAXS) revealed that M is highly structured and only moderately extended relative to N (51 Å vs 47 Å) (*85, 86*). Moreover, the comparison between the solvent protection and the chemical probing patterns of M and N revealed that they have almost the same secondary and tertiary structure, with the only differences distributed mainly in the core region such as P7 (*86*). Surprisingly, P3, instead of alt P3, was shown to be present in M. Nevertheless, it took hours to switch between these two highly similar structures. The refolding is accelerated by disrupting any of the five long-range tertiary contacts or by addition of urea, but it is impeded by elevated Mg^{2+} concentration. To reconcile all these data and the previous work, a topological model was developed by computational modeling and proposed that

alt P3 may bias a non-native topology error within the core that leads to M, but it reforms to native P3 in the final structure of M (Figure 1.7B) (*86*).

1.4.3 Lessons from other group I RNAs

1.4.3.1 Group I introns share a conserved core

More than 2000 sequences of group I introns have been classified by Michel and Westhof into five large groups and thirteen subgroups, according to the sequence in the conserved core and arrangement of peripheral elements (*73*). Despite some sequence variations, the core is conserved in the secondary structure (*73*) and three-dimensional structure, which has been demonstrated by crystallography for three group I introns belonging to different subgroups and organisms: group IC1 intron from the ciliate *thermophila* (*75*), group IC3 intron from the purple bacterium *Azoarcus* (*87, 88*), and group IA2 intron from the bactertiophage *Twort* (*89*). Cech compared these three structures and found that the core elements superimpose on each other with a small root mean square deviation (RMSD) (*16*) (Figure 1.8).

1.4.3.2 Stability of group I introns relies on the well-organized peripheral elements

A study surveyed twelve different group I introns from various subgroups and revealed that high GC content (≥35%) and optimal peripheral interactions secure the structural stability and self-splicing activity. Introns that fall out of these criteria require protein chaperones for activity (*90*). In consistent with this finding, the three crystallized introns mentioned above are stabilized by well-organized peripheral elements and exhibit robust activities on their own.

In the *Azoarcus* group I ribozyme, only two essential domains, P4-P6 and P3-P9, are present. They are covalently linked between P4 and P7 stems in the center, while each end of them is connected by a tetraloop-receptor interaction (Figure 1.9). One of the connections is L9-P5, and the other is L2-P8. The latter is important for the ribozyme activity since the reactive helix P1 that contains the substrate recognition site stacks on P2 and thus is positioned by the docking between L2 and P8. This small ribozyme forms a very stable global structure with each end forming by a tetraloop-receptor interaction. This optimized rearrangement and the abnormally high GC content (71%) agrees with the steady activity of the *Azoarcus* ribozyme at high temperature (*91*).

Larger and more intricate than the *Azoarcus* group I RNA, *Twort* (GC content=38%) and *Tetrahymena* ribozyme (GC content=45%) still manage to stabilize their global structure and maintain activity at 50 °C or higher temperatures (*40, 89*). In the *Twort* structure, there are two extra elements outside the P3-P7 core, P9.1 and P7.1/P7.2, which form a base pairing tertiary interaction with one another at the end of each helix (Figure 1.9). Apparently these extra elements did not increase the flexibility of the overall structure. However, the *Tetrahymena* ribozyme uses quite different strategy compared to the other two introns. Its P2 stacks on P2.1 instead of P1. The end of P2 stem is stabilized by P5abc, a distinct element in this ribozyme; while the other end of this long stacked stem contacts the P9.1/P9.2 stem (Figure 1.9B). Peripheral elements wrap around the active site completely in the *Tetrahymena* ribozyme, but only halfway in the *Twort* ribozyme (*16*).

In contrast, there are several group I introns rely on the stabilization of proteins in order to fold into the correct tertiary structure. The *ND1* group I intron from *Neurospora*

crassa has a similar structure to *Azoarcus*, but with lower GC content level (36%). Its core is less wrapped because of the lacking of P2 helix and the tetraloop-receptor contact between L2 and P8. *In vitro* studies showed that this ribozyme is only functional in the presence of a homogenous protein CYT-18 (*92*). A crystal structure of CYT-18 bound ND1 intron showed that the protein cofactor binds to the L9-P5 tetraloop-receptor, indicating it may assist contact formation there (*93*). This difference between the *Azoarcus* and *ND1* introns also indicates the cooperativity between tertiary elements because the L9-P5 contact in the former is readily formed in the presence of a tertiary contact (L2/P8) far from it. The bI5 and bI3 group I introns from the yeast mitochondria must also be stabilized by the proteins CBP2 and bI3 maturase/Mrs1, respectively (*94, 95*). Both of them have poor GC content and redundant tertiary elements.

1.4.3.3 Misfolding of other group I introns

Alternative folding has been broadly found in various group I RNAs, emphasizing the common theme of misfolding in structured RNAs. The T4 phase-derived thymidylate synthase (td) group IA2 intron, belonging to the same subgroup as the *Twort* intron, was demonstrated to adopt a misfolded structure *in vivo* and *in vitro*. It has a longer J2/3 that was found to be the source of misfolding—this J2/3 competed with substrate to bind the 3' of internal guide sequence (*96*).

A group IE intron from *Candida albicans* has analogous secondary structure as the *Tetrahymena* ribozyme except that its P5abc element is relatively short. This P5abc only constitutes by a long helix that is separated by two small internal loops; therefore the contact between P2 and P5abc might be lost. Folding experiments showed that this

ribozyme also has a misfolded intermediate (*97*), which can be resolved by initial folding in monovalent ions like the *Tetrahymena* ribozyme (*98*). Furthermore, *in vitro* and *in vivo* data indicated that the failure of the formation of the native P3 pseudoknot leads to the accumulation of intermediates (*97, 99*). Though misfolding is prevalent in large group I RNAs, it is avoided by the smaller *Azoarcus* ribozyme under standard conditions (*100, 101*), suggesting that structural complexity and peripheral elements increase the chance of kinetic traps.

1.4.3.4 Novel group I RNAs extend the knowledge of RNA study

The well-studied *Tetrahymena* ribozyme has provided us numerous data on the structure and folding of RNA. However, more structured RNAs must be studied in depth to further develop our understanding of them. Systematic data from a few other structured RNAs, such as the *td* group I intron, *Candida* ribozyme and *Azoarcus* ribozyme, have already expanded our scope to a larger degree. A previous study has discovered group IC1 introns at position 516 in the small subunit 18S rRNAs of red algae (*Bangia*) (*102*). They possess all of the typical IC1 subgroup elements as seen in the *Tetrahymena* ribozyme, except for two novel features that distinguish them from the other group IC1 introns (*103*). The first one is a six base pair insertion in P5b that is capped by a GNRA tetra-loop. The other is a short P8 stem that only contains three base-pairs, which are capped with a GNRA tetra-loop; while other group IC1 introns have varied P8 lengths and lack the tetraloop. Because GNRA tetraloops that cap duplexes with fixed length always form tertiary interactions in large structured RNAs (Lee.J and Gutell.R, unpublished data), it is very likely that those two novel features indicate additional

contacts in *Bangia* group IC1 introns. Investigating the structure and folding of this novel *Bangia* group I intron, which shares a similar GC content and conserved core as the *Tetrahymena* ribozyme, will further illuminate the importance of peripheral elements and their arrangements.

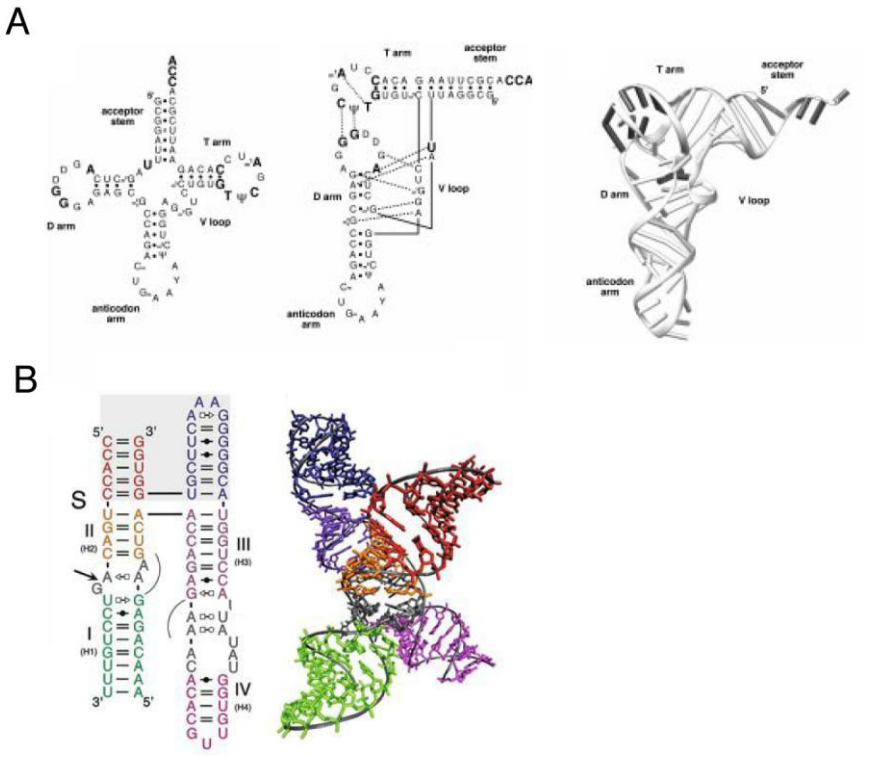

Figure 1.1: Structure of small RNAs

(A) Secondary and tertiary structured of tRNA. On the left is the secondary structure for the classic cloverleaf style. The middle is the rearranged secondary structure showing the tertiary interactions. The right picture is the 3D structure from crystallization (*104*). (B) Secondary and tertiary structure of the four-way junction hairpin ribozyme. Helices I and II, III and IV are coaxially stacked as shown in both the secondary and tertiary structure pictures (*105*).

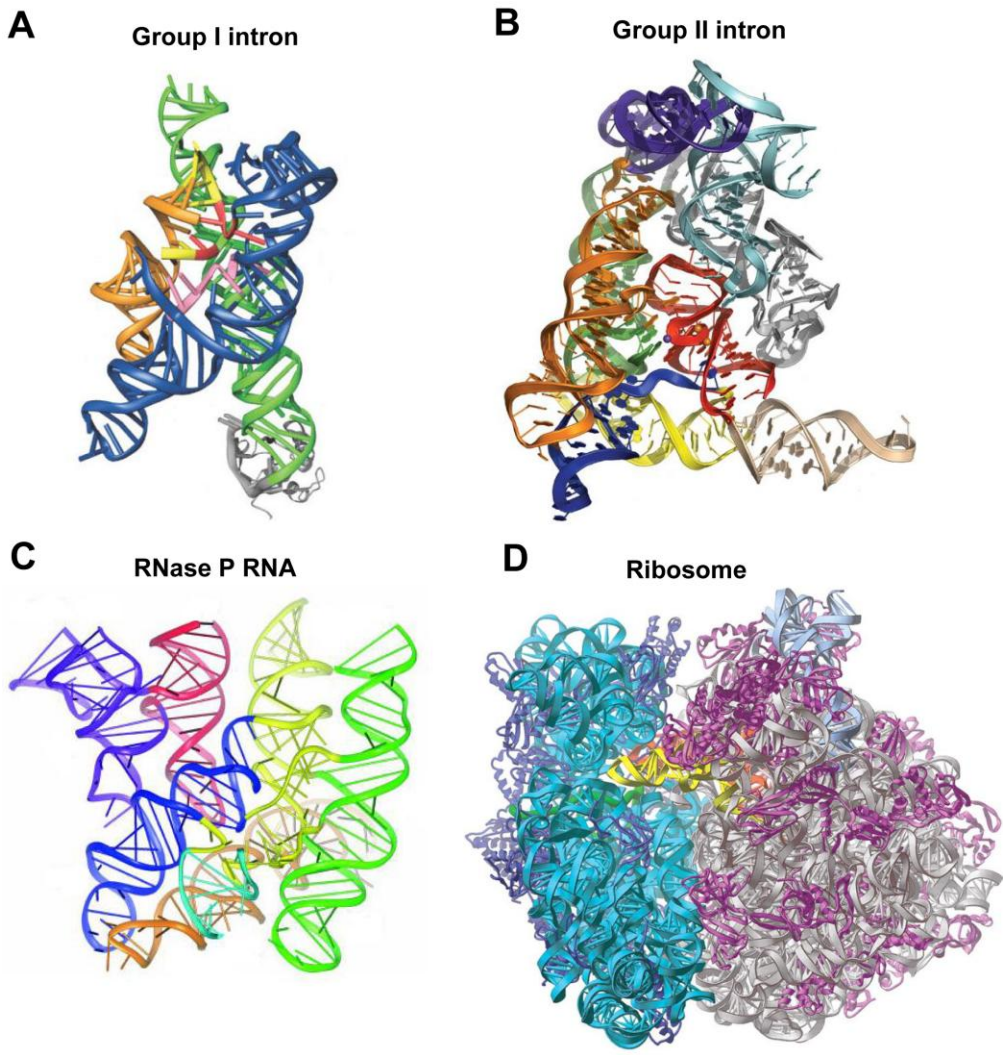

Figure 1.2: Tertiary structure of large catalytic RNAs/RNP

The crystal structures for different functional RNAs showed that helical domains are linked covalently or by tertiary motifs to form the global conformation. (A) Group I intron RNA from *Azoarcus* (*87, 88*); (B) Group II intron from *Oceanobacillus iheyensis* (*106*); (C) RNase P RNA from *Bacillus stearothermophilus* (*107*); (D) 70S ribosome (5S (gray-blue), 16S (cyan), 23S (gray) rRNAs) with protein components (purple and blue) from *Thermus thermophilus* (*23, 24*).

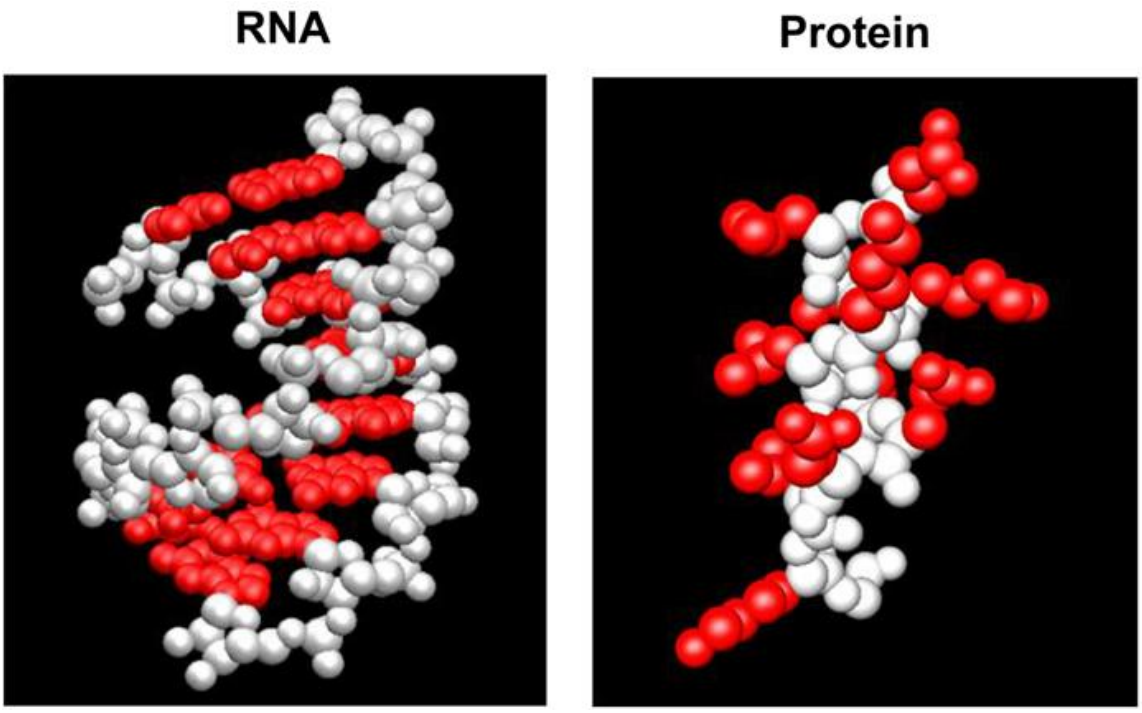

Figure 1.3: Difference between RNA and protein chains

The left panel shows an A-form RNA helix in which the phosphodiester backbone sugars are colored in white and the bases are in red. There are only 4 types of bases in RNA, and they are buried inside the helix. In the right panel, the protein polypeptide chain is colored so that the backbone is white and side chains are red. The side chains in protein chain point outwards to interact with each other.

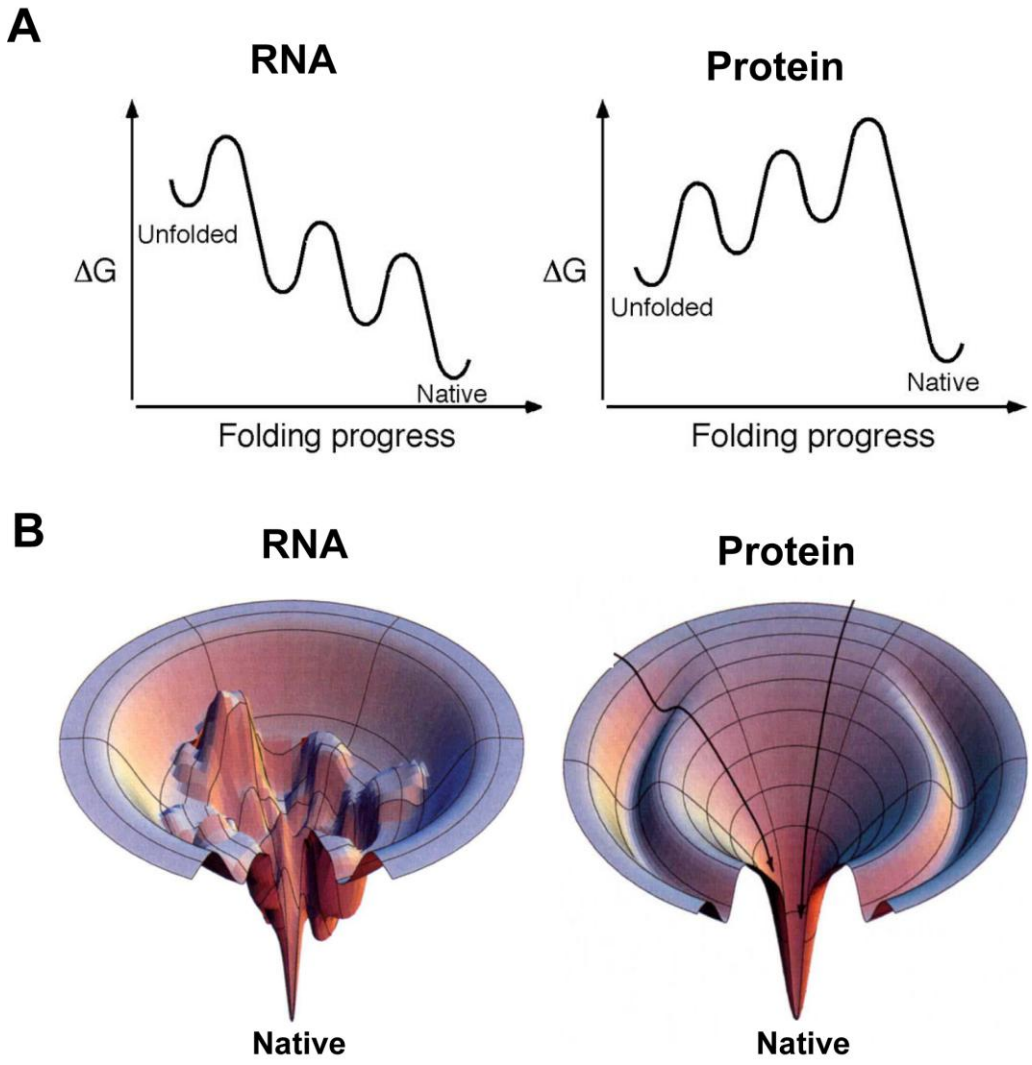

Figure 1.4: Folding landscaped of RNA and protein

(A) Folding profiles of RNA and protein on 2D scheme. The free energy of unfolded RNA descends downhill roughly towards the native state that has the minimum energy level. In contrast, many proteins, such as single-domain ones, undergo two-state folding. (B) Theoretical funnel landscapes for RNA and protein folding. Towards the native structure, unfolded RNA ensembles have to move down a rugged energy landscape with hills representing various intermediates to the native minima. The landscape of protein folding is smoother and involves none or few kinetic traps or energy barriers. (Adapted from reference (*8*).)

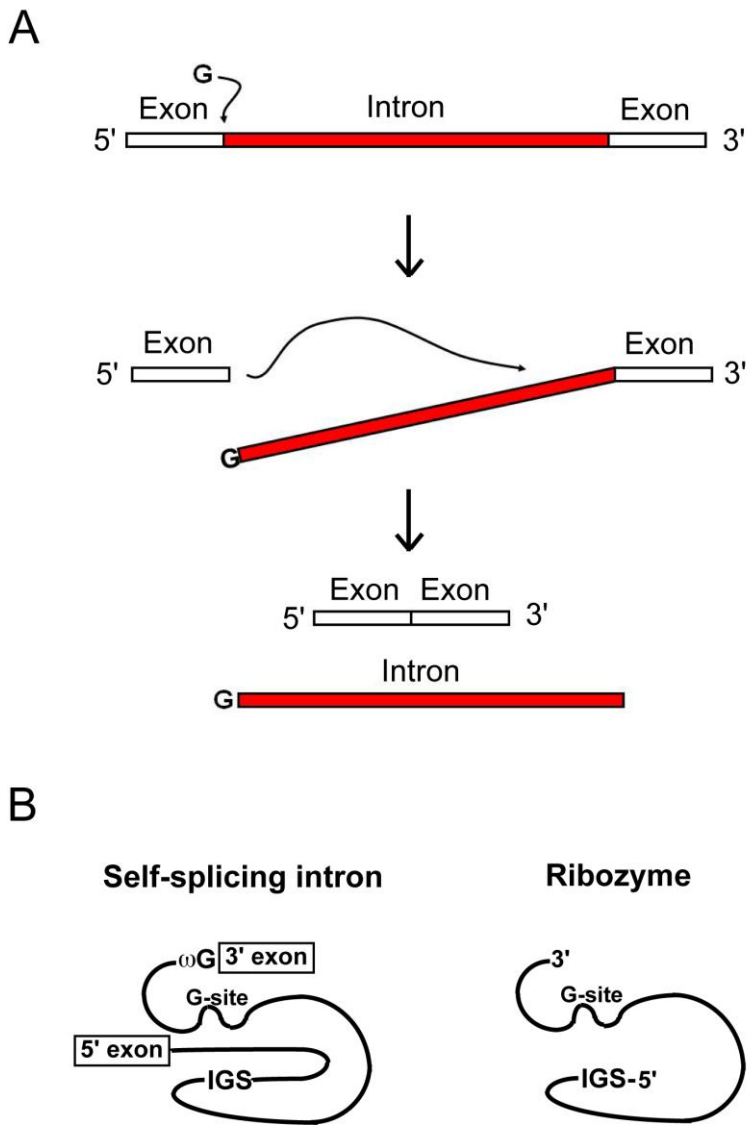

Figure 1.5: Self-splicing group I intron and its engineered ribozyme

(A) The two-step scheme of group I intron self-splicing. In the first step, the 3' hydroxyl of an exogenous guanosine attacks the 5' splicing site and links itself to the 5' of the intron. In the second step, the 3' hydroxyl of the 5' exon attacks the 3' splicing site, resulting in a free intron and ligated exons. (B) Engineering of group I intron into the ribozyme version. Both flanking exons and ωG are removed, leaving a ribozyme with an internal guide sequence (IGS) that recognizes the *trans* substrate at the 5' end (*108*).

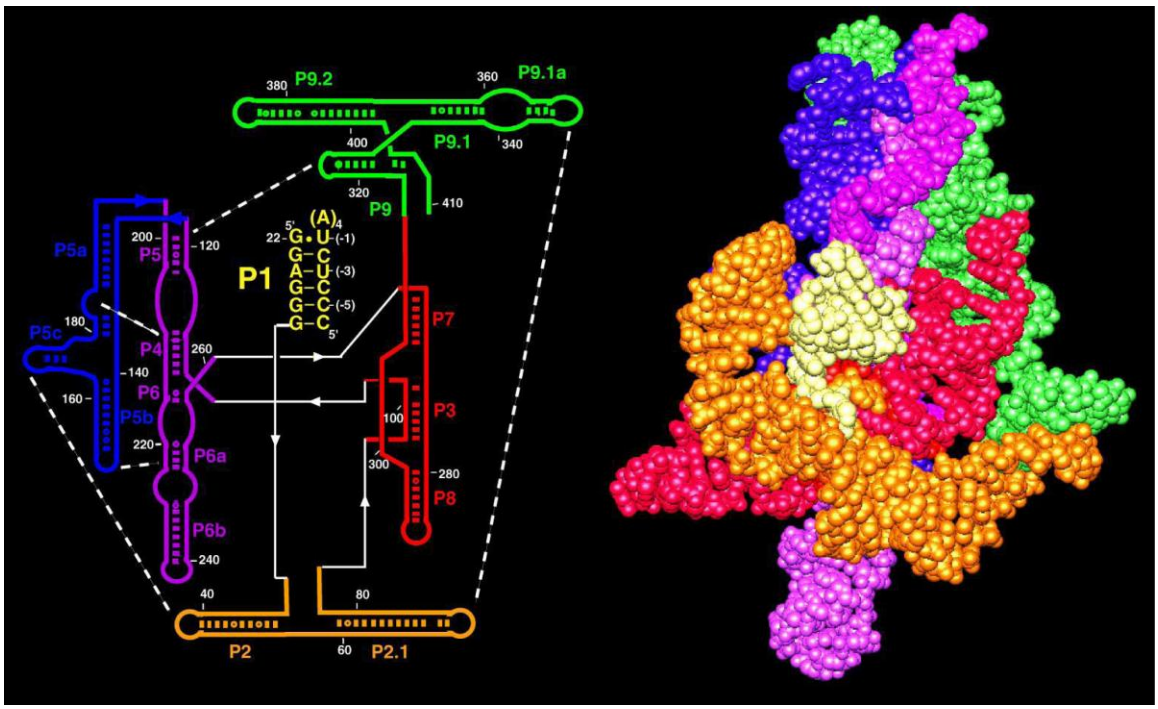

Figure 1.6: Structure of the model system: *Tetrahymena* group I ribozyme

The secondary structure on the left was colored according to the paired regions (P). The reactive helix P1 (yellow) and the core helical domain P4-P6 (purple) and P3-P8 (red) are surrounded by the peripheral helices P2-P2.1 (orange), P5abc (blue), and P9-P9.2 (green), which are connected by five long-range tertiary contacts (white dash lines). The three-dimensional structure on the right is adapted from a 3D model of the *Tetrahymena* ribozyme (*76*) that has been validated by biochemical data and crystallography from major domains (*75, 77, 78*). It is colored accordingly and shows the arrangement spatially.

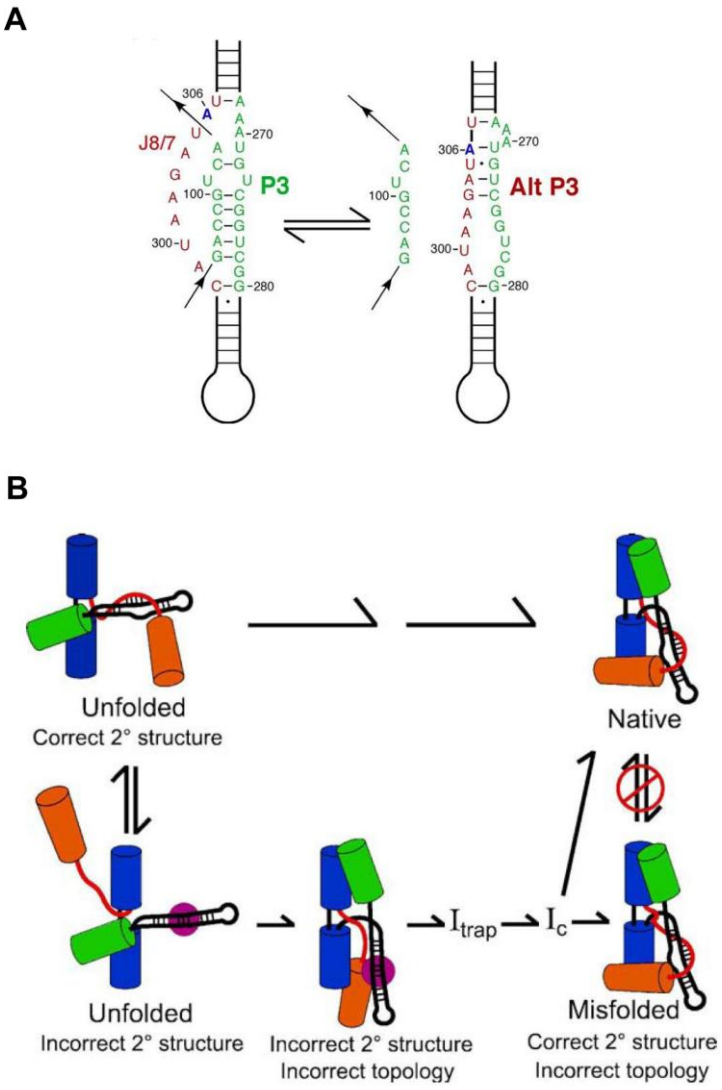

Figure 1.7: Model of Tetrahymena ribozyme folding pathway and alternative intermediate

(A) The secondary base-pairing of the native P3 pseudoknot and misfolded alt P3 (*38, 86*). (B) Model of misfolding for the *Tetrahymena* ribozyme due to topological error. In the early stages, unfolded RNA may either form the correct P3 structure that leads to formation of the native ribozyme directly (top panel) or a misfolded Alt-P3 structure that generates a topological error inside the core (bottom panel). This error results in a misfolded structure that is difficult to interconvert to the native ribozyme. However, alt P3 is replaced by P3 in this misfolded species eventually. Pictures are adapted from reference (*86*).

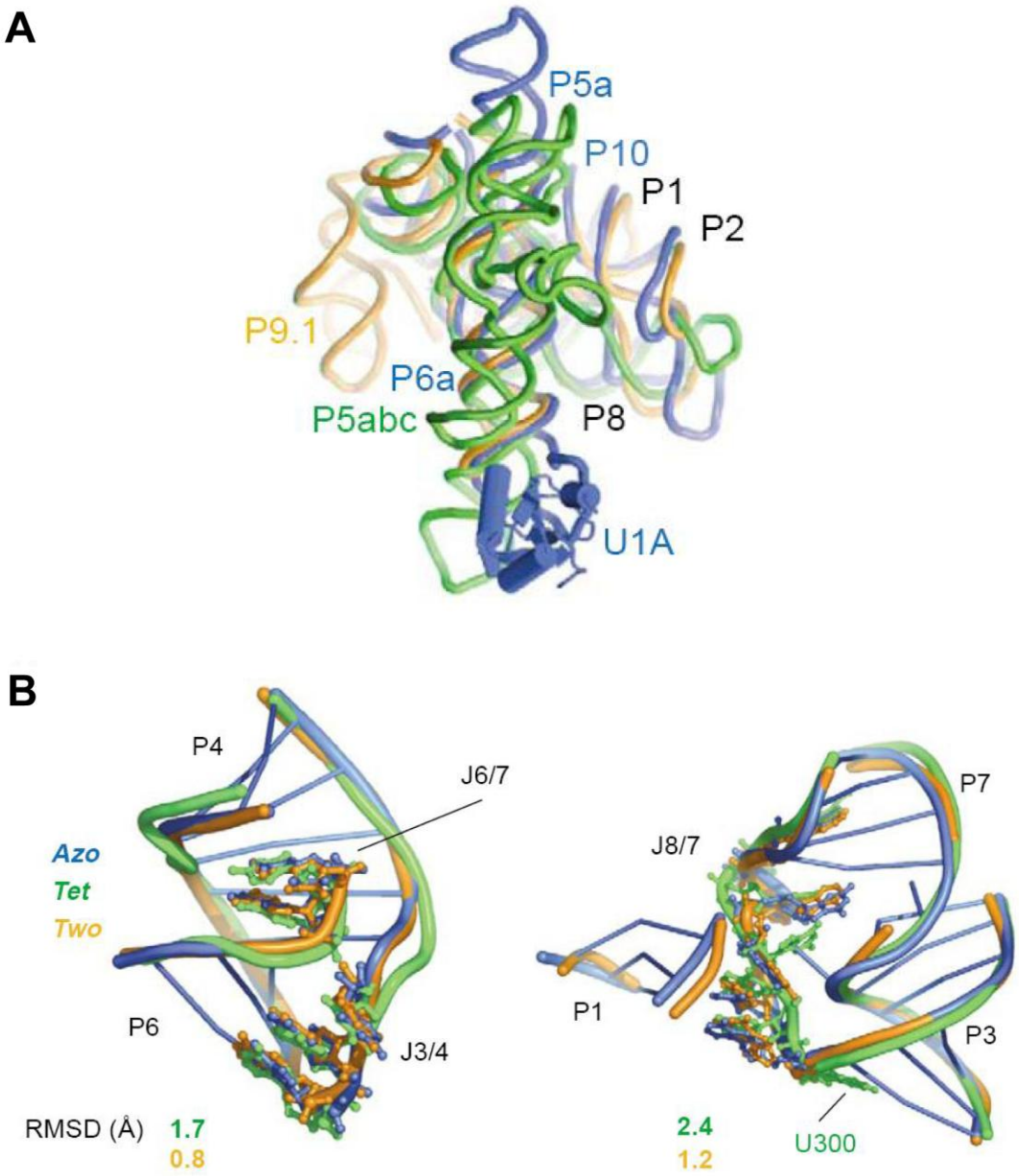

Figure 1.8: The overlapping of three group I intron crystal structures

(A) Superimposition of three group I introns: *Tetrahymena* (green), *Azoarcus* (blue) and *Twort* (orange) group I introns. (B) The enlarged core elements in the superimposition. P4-P6 (left panel) and P3-P7 (right panel) constitute the core of group I introns. Comparing *Tetrahymena* (green) and *Twort* ribozyme (orange) to the *Azoarcus* ribozyme (blue) gave RMSD values around 1-2 Å at both regions, indicating a highly conserved tertiary structure in the core. Pictures are from reference (*16*).

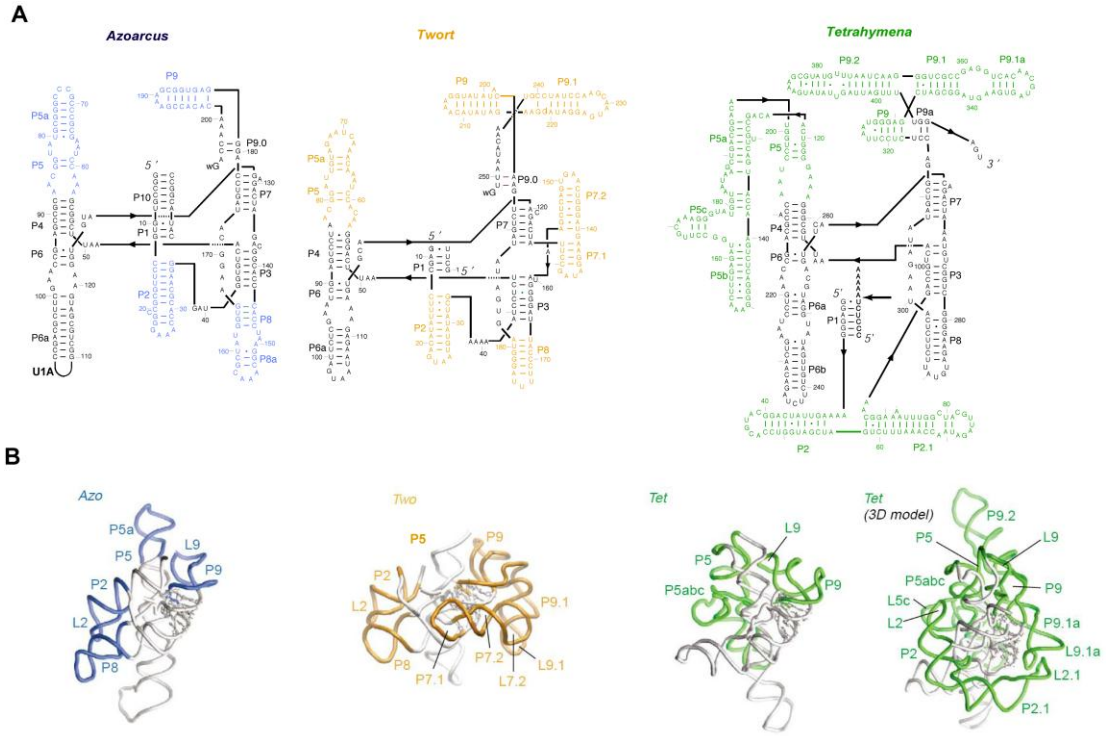

Figure 1.9: Arrangements of peripheral elements in group I introns to buttress the core

(A) Arrangement of group I intron peripheral elements in the secondary structure diagrams. The conserved cores are colored black in all three introns, while the peripheral helices are colored differently. (B) The tertiary structures of three group I introns. The conserved cores are colored in gray, and peripheral helices are highlighted accordingly. Picture is adapted from reference (*16*).

Chapter 2: Enhanced specificity against misfolding in a thermostable mutant of the *Tetrahymena* ribozyme

2.1 INTRODUCTION

To be functional, structured RNAs are faced with the challenge that they must selectively populate conformations representing a very small portion of the energy landscape. An important component of this challenge is to generate active conformations that are more stable than the most stable alternative ones, such that these functional structures remain populated at equilibrium (*7, 11, 109, 110*).

How this specificity is achieved during evolution is not clear. In general, two scenarios can be envisioned. In one, the RNA is under selective pressure to stabilize the native structure relative to partially folded or unfolded forms, and a sufficiently large fraction of mutations stabilize contacts that cannot form in any given alternative structure. Thus, selection for native stability alone produces enough specificity against all viable alternative structures. Alternatively, it is possible that negative selection is necessary. In this process, some mutations would become fixed in the RNA population not because they stabilize the native state relative to unfolded structures, but instead because they destabilize a particular family of misfolded structures that would otherwise be sufficiently populated as to interfere with the functions of the RNA. There is increasing evidence that some proteins have evolved in this latter scenario, with negative selection to reduce the ability to form extended beta sheet structures or other misfolded

structures (*111, 112*), and this principle has been incorporated into protein design efforts (*113-115*).

Although it is difficult to disentangle the complex and varied selective pressures experienced by RNA molecules in natural environments, artificial selection can be a powerful tool because it can define explicitly the selective pressures. An *in vitro* selection scheme was used previously to generate a mutant of the *Tetrahymena* group I intron ribozyme with substantially enhanced thermostability, as demonstrated by its ability to migrate rapidly at higher temperatures in temperature-gradient gel electrophoresis (TGGE) (*116*). This mutant has nine single-nucleotide substitutions that are localized predominantly away from helical elements, and further work showed that the effects of individual mutations are strikingly non-additive, suggesting that the enhanced stability arises largely from cooperative effects of the mutations (*117*). Analogously, an RNase P RNA of thermophilic origin was also shown to be more stable than its mesophilic counterpart and to display increased cooperativity in tertiary structure formation (*110, 118*).

In addition to forming a native structure, the *Tetrahymena* ribozyme folds *in vitro* to a long-lived misfolded intermediate under standard conditions (*9, 38, 39, 58, 81, 83, 86*). This misfolded conformation shares extensive similarity with the native structure, including a network of long-range peripheral contacts (*86*). Although it is defective for the standard ribozyme catalytic reaction, cleavage of an oligonucleotide that mimics the 5′ splice site, the misfolded ribozyme catalyzes efficient cleavage at the 3’-splice site, underscoring the similarity between the two structures (*86*). Despite this similarity, the native state is much more stable, with a robust energy gap of 6 kcal/mol (*119*).

The previous *in vitro* selection scheme included a step in each cycle that required the RNA to remain active, ensuring that the native state remained the most stable form. However, there was no selective pressure to increase or even to maintain the substantial energy gap between the native and misfolded structures. Here, we measure the relative stability of these two conformations for the evolved ribozyme mutant and find that the energy gap is increased by nearly 10-fold to >7 kcal/mol. Analysis of subsets of the mutations shows that the enhanced specificity against misfolding arises exclusively from two of the nine point mutations, which are close in space and cooperatively stabilize the native state (*117*). The environment of these nucleotides differs in the misfolded conformation (*86*), suggesting that the enhanced specificity against misfolding arose from fortuitous strengthening of a contact that cannot form in the misfolded conformation. Similar processes in nature may allow structured RNAs to achieve large energy gaps between their native and misfolded structures, even closely related ones, without explicit negative selection.

2.2 MATERIALS AND METHODS

2.2.1 Preparation of RNA

The R14C$^{\Delta P5abc}$ ribozyme was constructed by Quickchange PCR (Stratagene) from the gene encoding the full-length R14C using oligonucleotides encoding the 76-nucleotide P5abc deletion (*119*). Individual substitutions were also made using Quickchange, and the complete sequences of all mutant ribozymes were verified by DNA sequencing. Catalytic activity of all mutants was measured and found to be within 3-fold of the wild-type ribozyme (data not shown). Full-length ribozymes, P5abc-deleted

versions, and P5abc RNA were prepared by *in vitro* transcription using T7 RNA polymerase and purified by Qiagen RNeasy columns (*120*). Oligonucleotides (Dharmacon, Lafayette, CO) were 5′-end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and purified by non-denaturing polyacrylamide gel electrophoresis (PAGE) (*121*).

2.2.2 Catalytic activity assay to monitor native ribozyme formation

The refolding of misfolded ribozyme was followed by catalytic activity assay. P5abc-deleted ribozymes (200 nM) were folded to a mixture of the native and misfolded conformations by adding 10 mM Mg^{2+} at 25 ℃ (50 mM Na-MOPS, pH 7.0) (see Fig 2A). At various times thereafter, aliquots were mixed with a 'folding quench' solution containing guanosine and P5abc (600 nM final concentration). This solution was incubated for 5 min, sufficient time for P5abc to bind and activate the P5abc-deleted ribozymes for substrate cleavage but not enough time to allow interconversion of the native and misfolded conformers (*83, 86*). Refolding of full-length ribozymes was measured in the same way except that the folding quench did not include P5abc. Radiolabeled substrate CCCUCUA$_5$ (S*) was added and aliquots were quenched after 1-2 min by adding 2 volumes of 90% formamide, 20 mM EDTA (1 min for the wild-type ribozyme and 1.75 min for mutants with modestly reduced cleavage rates). Labeled substrate and product were separated by 20% denaturing PAGE and quantitated by a phosphorimager. The fraction of native ribozyme was determined by the fraction of S* that was cleaved to the shorter product CCCUCU (P*) (*83, 84, 122*). To correct for a small fraction of damaged and inactive ribozyme (<20%), data are shown after

normalization against an equivalent reaction in which the ribozyme was incubated at 50 ℃ in the presence of P5abc for 30 min to give essentially 100% native ribozyme (*83*). For P5abc-deleted ribozymes, observed rate constants for the approach to equilibrium (k_{obs}) were converted to rate constants for refolding from M to N ($k_{M-->N}$) by using the formula $k_{M-->N} = k_{obs}$ ($K_{M<-->N}/K_{M<-->N} + 1$), where $K_{M<-->N}$ is the equilibrium between the native and misfolded forms.

2.2.3 Mg^{2+} dependence of native ribozyme formation

The Mg^{2+} dependence of native ribozyme formation was measured as described previously (*123*). Briefly, the ribozyme or its variants were incubated with varying Mg^{2+} concentrations (0–1 mM for full-length ribozymes and 0–20 mM for P5abc-deleted ribozymes) at 25 ℃ (50 mM Na-MOPS, pH 7.0) for 2 h to allow equilibration of folded and unfolded conformers (see Fig 5A). For $E^{\Delta P5abc}$, P5abc (500 nM) and additional Mg^{2+} (50 mM) and guanosine (500 μM) were then added to restore catalytic activity of ribozyme that was in the native state in the initial incubation, whereas the majority of the ribozyme that was non-native was subsequently trapped in the misfolded conformation. After 10 min to allow full binding of P5abc, labeled substrate was added to initiate reactions. Substrate cleavage reactions with full-length ribozyme were performed similarly, except that Mg^{2+} and S were added, followed by guanosine to initiate the cleavage reaction. Reactions were quenched after 1 min for wild-type ribozyme or after 1.75 min for R14C ribozyme variants, sufficient time for complete substrate cleavage by the native ribozyme. To compare visually the Mg^{2+} dependences of different ribozyme variants, the results were normalized by subtracting the small fraction that avoided

misfolding after addition of 50 mM Mg^{2+} and dividing by the maximal fraction of native ribozyme. For example, the $R14C^{\Delta P5abc}$ ribozyme began at a higher value and progressed to a higher endpoint value, the former reflecting a more favorable kinetic partitioning during the folding process and the latter reflecting the increased specificity for folding to the native state (0.91 and 0.56 for $R14C^{\Delta P5abc}$ and $E^{\Delta P5abc}$, respectively).

2.2.4 P5abc binding and dissociation

To measure dissociation of P5abc, wild-type or mutant ribozymes (100 nM) were incubated for 15 s (50 mM Na-MOPS, 10 mM Mg^{2+}, 25 ℃) to generate a mixture favoring the misfolded ribozyme or for 1 h to generate predominantly native ribozyme. A trace amount of radiolabeled P5abc was added for 8 min to allow full binding, and then excess unlabeled P5abc (1 μM) was added. The fraction of labeled P5abc remaining bound over time was determined by loading aliquots on a continuously-running 12% native PAGE at 4 ℃ and analyzed by using a phosphorimager.

P5abc binding kinetics were measured analogously. The ribozyme was incubated as above to generate populations of predominantly native ribozyme or as much misfolded ribozyme as possible, and then labeled P5abc was added. At various times thereafter, excess unlabeled P5abc (2 μM) was added to block further binding of labeled P5abc, and aliquots were processed as above. Rate constants for binding to the native and misfolded conformations were calculated from the observed rate constants with these two mixtures as described previously (*119*).

2.2.5 Calculation of native specificity in the full-length ribozyme

The equilibrium values relating the stabilities of the native and misfolded conformations for the R14C ribozyme and other variants were calculated as described previously (*119*). Briefly, the corresponding equilibrium value for each P5abc-deleted ribozyme was multiplied by the affinity difference of P5abc for the native vs the misfolded core. This analysis makes the assumption that the behavior of the *trans* complex with P5abc is the same as that of the corresponding full-length ribozyme, as suggested by equivalent catalytic activity and structure probing results (*119, 124, 125*). Because the equilibrium values for the full-length ribozymes are calculated from experiments in which incubations were chosen to give extensive population of the native or misfolded conformation as desired, the values are not subject to errors that could arise from identifying exceedingly minor populations (*i.e.* establishing that 1 in 10^5 molecules are misfolded rather than being non-functional as a result of damage).

2.2.6 Hydroxyl radical footprinting using Fe (II)-EDTA

Before the footprinting assay, R14C ribozyme was 5′-end-labeled with [γ-^{32}P]ATP and 3′-end-labeled with [α-^{32}P]ATP as described below. To 5' end-label the R14C ribozyme, the RNA was first incubated with alkaline phosphatase (Fermentas) at 37°C for an hour to remove its phosphoryl group from the 5'end (*126*). The phosphatase was inactive afterwards by incubating at 65°C for 15 minutes. The ribozyme was then incubated with [γ-^{32}P] ATP (Perkin Elmer, 5 mCi stock) and T4 polnucleotide kinase (Fermentas) at 37°C for another hour and quenched by 20 mM EDTA, 0.01% bromophenol blue and 0.01% xylene cyanol in 95% formamide. To 3' end-label the

ribozyme, complementary DNA oligonucleotide to the 3' end of the ribozyme with a two nucleotide overhang and a 2'3' dideoxy C was designed as described in reference (*127*). The ribozyme (50 μM) was incubated with the DNA oligonucleotide (25 μM) and 1X annealing buffer (14 mM Tris-HCl pH 7.5, 40 mM NaCl and 0.2 mM EDTA) at 65°C for 1 minute and transferred to room temperature. Klenow fragment (500 U/ml) (New England Biolabs), [α-^{32}P] ATP (20 μCi/μl) and 1X Mg^{2+} buffer (7 mM $MgCl_2$, 1 mM DTT) were mixed with the reaction system at 37°C for 2 hours to add the radio-labeled ATP at the 3'end of the ribozyme. The volume of the final reaction system was concentrated from 100 μl to 20 μl using Micron YM-50 (Millipore) and purified using P-30 Bio-Spin column (Bio-Rad). The same quench buffer used in the 5'end labeling was then added to stop the reaction. Both 5' and 3' radio-labeled products were purified by running on the 8% native PAGE at 15 Watts for 1 hour. RNAs were excised and eluted in TE buffer (10 mM Tris-HCl, pH8.0, and 1 mM EDTA) for at least 8 hours at 4°C and stored at -20 °C.

The radio-labeled R14C ribozyme was pre-folded under standard folding buffer (pH 7.0, 50 mM Na-MOPS and 10 mM $MgCl_2$) at 50°C for 30 minutes to form majority of the native conformation or at 25°C for 1 minute to form about 60% of the misfolded conformation. Fe(II)-EDTA footprinting reagent (2 mM $(NH_4)_2Fe(II)(SO_4)_2$, 2.5 mM Na-EDTA, 6mM sodium ascorbate) with a 1/10 volume was mixed with the pre-folded ribozyme for 10 min at 25°C. Reactions were stopped in one half reaction volume of quench buffer (33 mM thiourea, 30% formamide, 0.03% bromophenol blue and 0.03% xylene). Products were resolved on an 8% denaturing gel at 55 Watts for 1-3 hours. The gel running buffer was the TTE buffer (70 mM Tris, 20mM taurine, 0.4 mM Na-EDTA).

RNase T1 ladder that identifies all the guanosines in the RNA sequence was generated by digesting the RNA with T1 endonuclease at 7M urea and ran with the footprinting samples. The intensity of bands in each reaction was analyzed by SAFA and normalized by the average intensity of the entire lane.

2.2.7 Thermal denaturation

RNA denaturation was followed on a DU7400 Beckman-Coulter UV spectrophotometer (260 nM). Ribozyme was incubated in 50 mM Na-MOPS, pH 7.0 (25 ℃), 5 mM $MgCl_2$ (80 ℃, 3 min), slowly cooled to room temperature and incubated for 5 min. Absorbance data were collected as the ribozyme was heated from 25 ℃ to 90 ℃ (0.5 ℃ per min).

2.3 RESULTS

2.3.1 Increased native specificity in the thermostable R14C mutant

To measure the relative stability of the native and misfolded conformations of the thermostable ribozyme mutant, termed R14C (ref. (*116*), we used a thermodynamic cycle developed previously with the wild-type ribozyme (*119*) (Figure 2.1). The earlier work showed that the equilibrium between the native and misfolded conformations for a version of the ribozyme that lacks the peripheral element P5abc ($E^{\Delta P5abc}$) is near unity (K = 1.4), as determined directly by catalytic activity (*119*). P5abc was shown to bind to the native core nearly 10^5-fold tighter than to the misfolded core. Completion of the thermodynamic cycle led to a calculated value of approximately 10^5 for the equilibrium

between the native and misfolded conformations of the complex and presumably of the full-length ribozyme (*119, 125, 128, 129*).

To perform analogous measurements of the R14C ribozyme, we constructed a variant that includes all nine mutations and lacks P5abc ($R14C^{\Delta P5abc}$; Figure 2.2A&B). Below, we describe experiments using $R14C^{\Delta P5abc}$ and P5abc added separately to measure the values of this thermodynamic cycle, and then we dissect contributions from individual and pairwise mutations.

We first measured the equilibrium of the $R14C^{\Delta P5abc}$ ribozyme between the native and misfolded conformations using catalytic activity (*83, 84, 122*). To determine the fraction of native ribozyme during Mg^{2+}-induced folding, we added radio-labeled substrate ($CCCUCUA_5$, S*) to aliquots of a folding reaction and measured the fraction of S* that was cleaved rapidly (Figure 2.3A&B). As expected, much of the $R14C^{\Delta P5abc}$ ribozyme did not reach the native state on the time scale of seconds, although the fraction that avoided misfolding was larger than for the wild-type ribozyme (Figure 2.3C). Hydroxyl radical footprinting of the misfolded R14C ribozyme revealed structural signatures of the misfolded wild-type ribozyme (Figure 2.4), suggesting a common origin of misfolding. Unlike the wild-type $E^{\Delta P5abc}$ ribozyme, however, $R14C^{\Delta P5abc}$ then refolded to give predominantly native ribozyme (K_{eq} = 8.4, Figure 2.3C). We confirmed that this endpoint reflected an equilibrium by folding at higher Mg^{2+} concentration, which gave a further shift toward the native state as observed previously for the $E^{\Delta P5abc}$ ribozyme (*119*), and then diluting back to 10 mM Mg^{2+} (Figure 2.3D). Upon dilution, the native fraction decreased to the same endpoint as upon initial folding at 10 mM Mg^{2+}. Thus, compared to the wild-type $E^{\Delta P5abc}$ ribozyme, $R14C^{\Delta P5abc}$ displays increased specificity in folding,

shifting the equilibrium between the native and misfolded conformations by a factor of six.

With knowledge of the rate and equilibrium constants for interconversion of the native and misfolded conformations of R14C$^{\Delta P5abc}$, we used a pulse-chase native gel shift approach to determine the affinity of P5abc for each conformation by measuring the kinetics of binding and dissociation (*119*). To measure dissociation, radio-labeled P5abc was bound either to a population of predominantly native R14C$^{\Delta P5abc}$ ribozyme, formed by allowing equilibration, or to a population that was incubated briefly with Mg^{2+} to produce an approximately equal mixture of native and misfolded ribozyme (Figure 2.5A). Upon addition of excess unlabeled P5abc chase, this latter reaction gave a significant dissociation phase with a rate constant of 0.0072 min^{-1} (Figure 2.5B). This phase was much less prominent in the reaction with predominantly native ribozyme, indicating that the observed dissociation is from misfolded ribozyme. In contrast, dissociation of P5abc from the native ribozyme was extremely slow, as observed previously for the wild-type E$^{\Delta P5abc}$ ribozyme (*119*). Extrapolation from higher temperature (Figure 2.5C) gave a rate constant of 2.4×10^{-7} min^{-1} ($t_{1/2}$ = 5-6 yr). An analogous pulse-chase approach was used to measure the binding rate of P5abc. From two sets of binding reactions, one with largely native ribozyme and one with a mixture of native and misfolded ribozyme, the binding rate constants were determined to be 8.6×10^{6} M^{-1} min^{-1} and 2.8×10^{6} M^{-1} min^{-1} for the native and misfolded ribozyme respectively (Figure 2.5D).

The much slower dissociation, coupled with the modestly faster binding, indicates that P5abc binds to the native ribozyme 90,000-fold tighter than to the misfolded

conformation (Figure 2.1), the same difference within error as for the wild-type ribozyme (Figure 2.1 and ref. (*119*). Completing the thermodynamic cycle gives a K_{eq} value for the full-length R14C ribozyme of 8×10^5, about 7–8-fold larger than for the wild-type ribozyme. Strikingly, the *in vitro* selection produced a ribozyme that is *better* able to discriminate between the native and misfolded conformations despite the absence of selective pressure to increase or even to maintain this energy gap.

2.3.2 The peripheral element P5abc is required for increased stability but not specificity

Insight into the origin of the enhanced specificity was obtained from further dissection of the effects of the R14C mutations in the presence and absence of P5abc. The results above indicated that the increase in specificity does not depend on P5abc, as it is the same for the P5abc-deleted and full-length ribozymes. On the other hand, the data suggested that the increased *stability* of the native conformation, relative to less structured or unfolded intermediates, requires P5abc. This model is illustrated in Figure 2.6A and B**,** in which the reference state is defined as being the transition state ensemble for refolding. This transition state reflects substantially unfolded conformations, as earlier work showed strong effects of urea and tertiary contact disruptions on the refolding rate (*86*, *123*). This behavior is likely to hold for the R14C ribozymes, because they are accelerated for refolding with the same urea dependences as the corresponding wild-type ribozymes (Figure 2.7A&B).

In the absence of P5abc, the R14C mutations increase the rate of refolding from the misfolded conformation (see Figure 2.3C), indicating a *decrease* of 1.7 kcal/mol in the

stability of the misfolded conformation relative to the reference state (Figure 2.6A). Although the energy gap between the native and misfolded conformers is increased for R14C$^{\Delta P5abc}$, this increase is smaller than the destabilization of the misfolded conformer, such that the native state is also modestly destabilized relative to this reference state. In contrast, in the full-length ribozyme, the R14C mutations reduce the refolding rate by 5-fold (Figure 2.8), indicating that the misfolded conformation is stabilized relative to the transition state reference (Figure 2.6B). The enhanced specificity for native folding indicates that the native state is stabilized to an even greater extent.

To test the idea that the enhanced stability requires P5abc, we set out to compare the stabilities of the native state for the R14C$^{\Delta P5abc}$ and E$^{\Delta P5abc}$ ribozymes. Surprisingly, temperature gradient gel electrophoresis (TGGE) did not reveal transitions at substantially lower temperatures than the full-length ribozyme versions, and UV absorbance measurements gave similar dependences, with transitions for both the wild-type and variant P5abc-deleted ribozymes at only slightly lower temperatures than the corresponding full-length ribozymes (data not shown). The transitions for the P5abc-deleted ribozymes were much higher than expected from previous work (*123, 125*), suggesting that they reflect principally the loss of secondary structure rather than functional tertiary structure. The corresponding transitions both occur at relatively high temperatures in the wild-type ribozyme (*130*), but they presumably become much less closely coupled in the P5abc-deleted mutants because the tertiary structure is destabilized substantially (*123, 125*).

We therefore turned to the Mg^{2+} dependence, with catalytic activity used as a readout for native state formation (*84, 122, 123*). It was shown previously that the Mg^{2+}

dependence for native ribozyme formation of $E^{\Delta P5abc}$ mirrors that for formation of tertiary structure monitored by hydroxyl radical footprinting (*123*). Using catalytic activity, we found that the $E^{\Delta P5abc}$ and $R14C^{\Delta P5abc}$ ribozymes gave indistinguishable Mg^{2+} dependences (Figure 2.9A&B), supporting the view that the mutations do not stabilize the native $E^{\Delta P5abc}$ structure relative to intermediates that lack stable tertiary structure. As expected, the full-length R14C required lower Mg^{2+} concentration than the wild-type ribozyme (Figure 2.9C). Apparently the net stabilization of the native state by the R14C mutations is entirely dependent on the presence of P5abc, despite the fact that none of the mutations are in P5abc. This striking result confirms and extends the previous conclusion that much of the stabilization arises from cooperative effects (*117*).

2.3.3 Two mutations are responsible for the increased specificity

The finding that the enhanced specificity for native state formation does not depend on P5abc suggested that it has a local origin, in contrast to the largely global origin of the increased stability. This idea raised the possibility that one or a small subset of the mutations underlies the enhanced specificity. We identified A269G and A304G as candidates because these mutations were shown previously to stabilize the native state when made in tandem (*117*) and because structure probing indicated that the limited differences between the native and misfolded conformations are concentrated in this region (*86*).

Therefore, we generated an $E^{\Delta P5abc}$ variant with the A269G and A304G substitutions and determined values for the thermodynamic cycle as above (Figure 2.10). In the absence of added P5abc, this ribozyme displayed enhanced specificity, with an

equilibrium value indistinguishable from that of R14C$^{\Delta P5abc}$ (K_{eq} = 6.8, Figure 2.11, bottom left). This enhancement required both mutations, with the individual substitutions giving smaller effects (Table 2.1, ribozymes 9 and 10). The preferential binding of P5abc to the native form was maintained in the mutant (Figure 2.10), with an affinity difference of 3×10^5-fold. Thus, completion of the thermodynamic cycle indicates that the A269G and A304G mutations impart at least as much specificity to the full-length ribozyme as the entire set of R14C mutations (K_{eq} = 2.1×10^6, Figure 2.10). This conclusion was confirmed using a variant in which 269 and 304 were reverted to their wild-type identities in the background of the R14C sequence. This ribozyme displayed a complete loss of the enhanced specificity with bound P5abc, giving a value indistinguishable from the wild-type ribozyme (Figure 2.12).

Further insight into how these substitutions enhance specificity came from the refolding kinetics. The A269G/A304G mutant did not display the increased refolding rate of R14C$^{\Delta P5abc}$, instead giving a value similar to that of E$^{\Delta P5abc}$ (Figure 2.11, bottom left). This result indicated that A269G and A304G have little or no effect on the stability of the misfolded conformation relative to the unstructured transition state for refolding. However, they do stabilize the native state, and we confirmed this conclusion by demonstrating a reduced Mg^{2+} requirement for native state formation in this mutant ($K_{1/2}$ = 0.86 mM, whereas wild-type ribozyme is 2.4mM, data not shown). Together these results suggest that 269G and 304G stabilize the native state by participating in contacts that are unable to form in the misfolded conformation (see Section 2.4. Discussion).

2.3.4 Dissection of independent energetic effects in the absence of P5abc

Although the mutations in A269 and A304 gave a clear enhancement in specificity in both the wild-type and R14C background, most of the effects of the R14C mutations were shown previously to be highly context-dependent (*117*). However, as described below, we found that in the absence of P5abc most of the mutations have no effect on stability or specificity, and the effects that do arise from mutations or sets of mutations are energetically independent, allowing for facile dissection.

The first hint for independent effects was that the A269G/A304G mutations increased the specificity of the full set of mutations but did not accelerate refolding of the misfolded conformer (Figure 2.3C). This result implied that a different mutation or set of mutations is responsible for the observed destabilization of the misfolded R14C$^{\Delta P5abc}$ ribozyme (Figure 2.6A). A candidate for this destabilization was U277C, which disrupts a base pair within P3 (Figure 2.2A). This mutation would be expected to destabilize the native and misfolded conformations because previous work suggested that P3 is formed in both conformations (*86*) but not in the unfolded ensemble (*50, 131*). Uniform destabilization could quantitatively account for the results described in Fig 3A if similar in magnitude to the native stabilization from A269G and A304G and provided that the effects were energetically additive.

Consistent with this hypothesis, a U277C variant of E$^{\Delta P5abc}$ destabilized the native and misfolded states equally, giving a refolding rate equal to that of the R14C$^{\Delta P5abc}$ ribozyme without shifting the equilibrium toward the native state (Figure 2.11, top right). Together, these results suggested a simple model in which A269G and A304G specifically stabilize the native state, U277C destabilizes the native and misfolded

conformations equally, and for the P5abc-deleted ribozyme, the mutations act independently. Indeed, adding together the energetic effects of these three mutations gives predicted behavior that is similar to the observed properties of the R14C$^{\Delta P5abc}$ ribozyme (Figure 2.11, bottom right). Further, a triple mutant (A269G/A304G/U277C) displayed the native specificity associated with the A269G and A304G mutations, the faster refolding rate associated with U277C, and gave results similar to the R14C$^{\Delta P5abc}$ ribozyme (Figure 2.11, bottom right).

To confirm this conclusion and to explore potential effects of sequence context, we also constructed a series of mutants starting from the wild-type nucleotides at positions 269, 304, and 277 but with the R14C mutations at the other six positions (Figure 2.13; Table 2.1, ribozyme 8**).** This mutant gave folding rates and equilibrium for interconversion of the native and misfolded species that are the same within error as the wild type, again indicating that the other six mutations have little or no effect on stability or specificity in the absence of P5abc. Further, against this background the A269G/A304G and U277C mutations gave the same effects as in the wild-type background.

Two important conclusions arise from these results. First, only three of the nine mutations give significant changes to the stability or specificity of the ribozyme in the absence of P5abc. The other six nucleotides most likely do not form base-specific tertiary contacts in the absence of P5abc and its tertiary connections (*132*). Second, the mutations that do give effects do so in an energetically independent manner, most likely reflecting a loss of global cooperativity in the absence of P5abc. Indeed, even the effects of mutations

in nucleotides 269 and 304 are not fully cooperative **(**Table 2.1, ribozymes 9 and 10**)**, in contrast to the effects in the full-length ribozyme.

2.3.5 Specific native state stabilization by 269G/304G in the full-length ribozyme

The results above suggested a simple model in which G269 and G304 interact energetically in the full-length ribozyme to increase the stability of the native state, relative to adenosines at these positions, but they do not form equivalent interactions to stabilize the misfolded state. Although the results above demonstrate that these mutations provide increased specificity in the full-length context, they do not address whether the mutations affect the stability of the misfolded conformation relative to less structured conformations. Therefore, we tested a A269G/A304G mutant in the full-length wild-type background and a G269A/G304A reversion mutant in the R14C background (Table 2.2, ribozymes 13 and 14). For both of these mutants, refolding from the misfolded conformation gave rate constants within 2-fold of those of the parental ribozyme (Figure 2.8), indicating that these substitutions have little or no effect on the stability of the misfolded conformation. Thus, this conclusion holds in the full-length ribozyme as well as in the P5abc-deleted version.

On the other hand, considerable context-dependent effects in the full-length ribozyme were noted previously (*117*), and we observe context dependence of some substitutions. In the wild-type background, U277C does not produce acceleration of refolding as it does in the $E^{\Delta P5abc}$ background (Table 2.2, ribozyme 15), perhaps because the loss of the base pair can be compensated by a tertiary contact (*117*). Nevertheless, the U277C substitution does produce a modest acceleration in the R14C background (2–3-

fold; Table 2.2, ribozyme 16). Thus, it remains possible that some mutations stabilize or destabilize the native and misfolded conformations differentially in certain contexts. However, because the mutations in A269 and A304 elicit the full increase in specificity in both the wild-type and R14C backgrounds, the other mutations produce no net change.

2.4 DISCUSSION

In the *in vitro* selection experiment that led to the generation of the R14C *Tetrahymena* ribozyme, the selective pressure was for enhanced thermostability, specifically the ability to migrate rapidly through a gel at increased temperature (*116*). Because the native state was the dominant state at the start of the selection and an activity check within the selection cycle ensured that it remained dominant, it was this state on which the selective pressure for stability was exerted. However, there was no explicit selective pressure to increase or even to maintain the large energy gap between the native and misfolded conformations. Nevertheless, our results indicate that the energy gap increased during the selection, with the R14C variant being better able to discriminate between the native and misfolded conformations than the wild-type ribozyme. This striking result indicates that structured RNAs are capable of evolving very large energy gaps between their native states and alternative structures, even closely-related structures, without negative selection against the alternative structures.

A further dissection of the effects of individual and pair-wise mutations has given physical insight into how the native state was stabilized relative to the misfolded conformation. The selective stabilization arose from just two of the nine mutations, A269G and A304G. These two mutations alone give the same increase in specificity as

the full set of nine mutations, in both the backgrounds of the wild-type and the R14C ribozymes. These two nucleotides are close to each other in the native structure (*75, 88*) and the mutations to G were shown previously to stabilize the native state, presumably by stabilizing a network of contacts involving both nucleotides (*117*). Although the misfolded and native conformations share a high degree of global similarity, structure-probing results indicate local differences that include and surround these nucleotides (Figure 2.4) (*86, 133*). Thus, the simplest model is that the connection that is formed or strengthened by the mutations is unable to form in the misfolded structure (Figure 2.14), most likely because local differences between the native and misfolded conformations result in re-positioning of one or both structural elements such that they do not contact each other.

Because the selective pressure was solely to stabilize the native state relative to unstructured forms, there was presumably no particular benefit from stabilization of the native state relative to the misfolded conformation. Indeed, current and prior evidence indicates that the other seven mutations stabilized both the native and misfolded conformations, which is not surprising given their similarity. The mutant that lacks the R14C mutations at positions 269 and 304 but has the other seven mutations refolds from the misfolded conformer 5-fold slower than the wild type, indicating that the mutations stabilize the misfolded conformer by 1 kcal/mol relative to the transition state. This mutant displays the same equilibrium between the native and misfolded conformers as the wild type, implying that these mutations stabilize the native and misfolded states equally in the wild-type background (Table 2.2, ribozyme 14). In the R14C background, the other seven mutations appear to stabilize the misfolded form with a small preference

(Table 2.2, ribozymes 12), consistent with prior results showing a modest effect of these mutations on stability of the native R14C ribozyme (*117*). The requirement for P5abc for enhanced stability from these mutations suggests that their contacts are only formed stably when the global architecture is stabilized, most likely from cooperative formation of the long-range contacts from P5abc (*132*), and these additional contacts would then be expected to contribute to the overall level of cooperativity (*117*). In contrast, the substitutions at positions 269 and 304 fortuitously stabilize a connection that cannot be formed in the misfolded conformation, and therefore they stabilize the native state specifically.

Previous results have suggested that structured RNAs can achieve enhanced stability through evolution by stabilizing networks of tertiary connections that form cooperatively (*110, 117, 118*), and our results suggest that these processes can generate substantial specificity against alternative structures, without explicit negative selection, because the alternative structures do not have all of the structural elements positioned to form the same networks of contacts. A key factor may be the relative rigidity of the helical elements that come together to form structured RNAs, which may prevent small-scale rearrangements and thereby enforce local differences between conformations (*120*). Rigidity is also likely to contribute by reducing the entropic penalties for fixation of helical elements, allowing even a single hydrogen bond to provide measureable stabilization relative to conformations in which the partners presumably form hydrogen bonds with water (*134, 135*). The presence of a large and robust energy gap between the native structures and inactive conformations has been suggested to be critical for ensuring that native RNA structures remain populated in the complex cellular environment (*110,*

136), and the results here demonstrate that this specificity can be evolved and maintained under selective pressure solely for thermostability of the native structure.

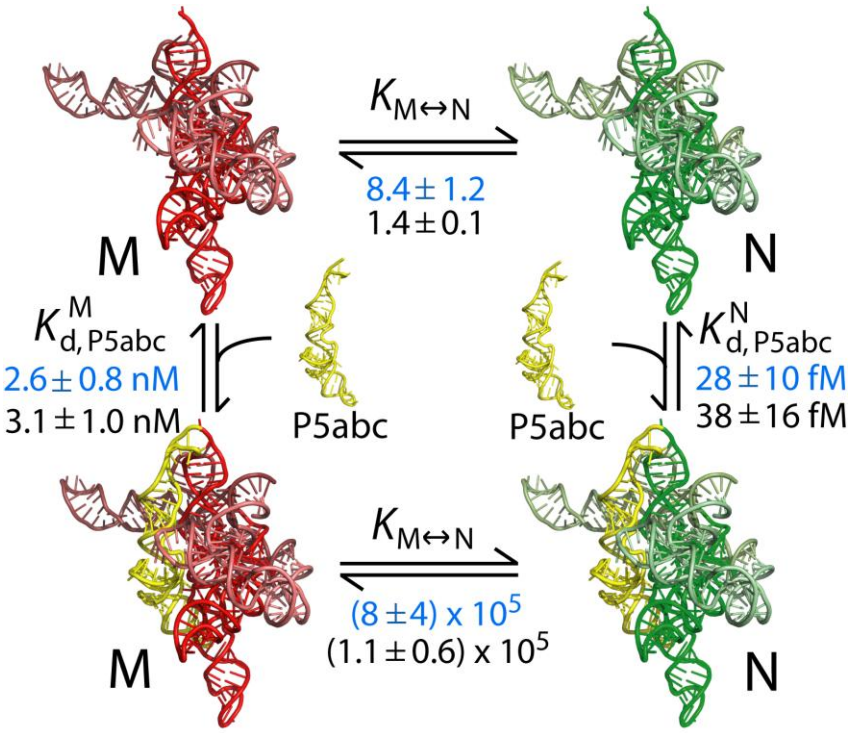

Figure 2.1: Thermodynamic cycle used to calculate the equilibrium between the native and misfolded forms

Here and in other figures, the native (N) and misfolded (M) conformations are depicted schematically, in green and red respectively, by showing the ribozyme structural model (*76*). Equilibrium constants obtained herein for the R14C ribozyme are blue and corresponding values for the wild-type ribozyme are black (*119*).

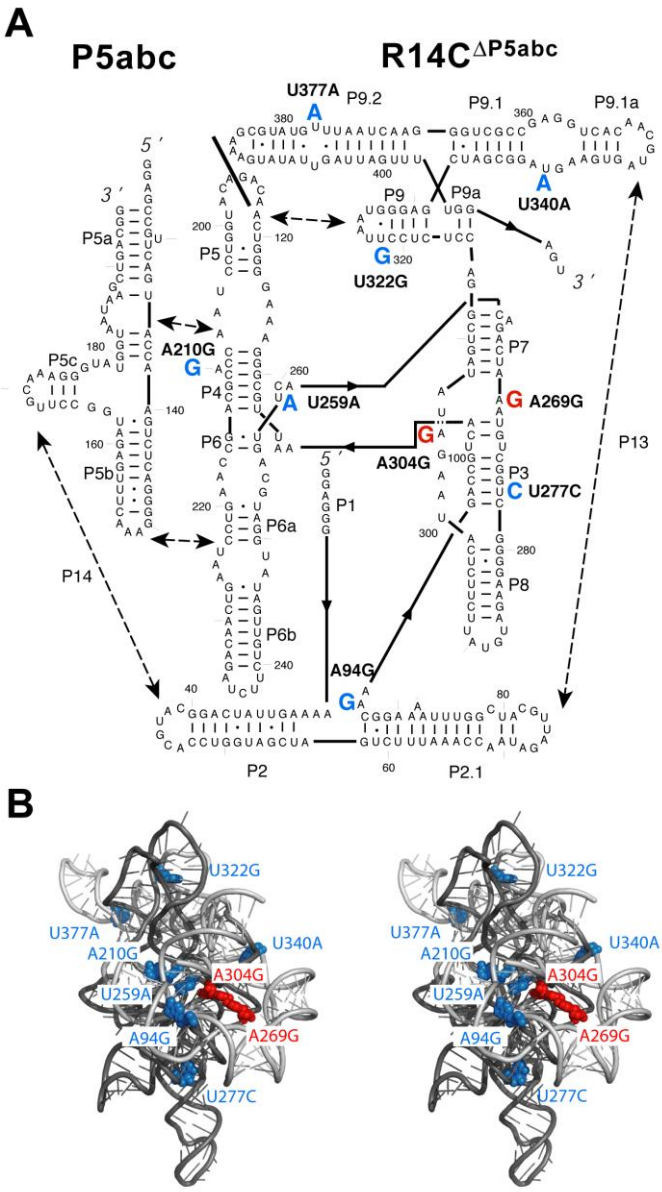

Figure 2.2: Secondary structure and tertiary structural model of the R14C ribozyme

(A) The two-piece ribozyme construct. The mutations in the R14C ribozyme are shown in large letters. The two substitutions shown herein to increase native specificity are red, and the other seven are blue. (B) Ribozyme stereo view showing the nine substitutions of the R14C ribozyme structural model. Colors are as in panel A.

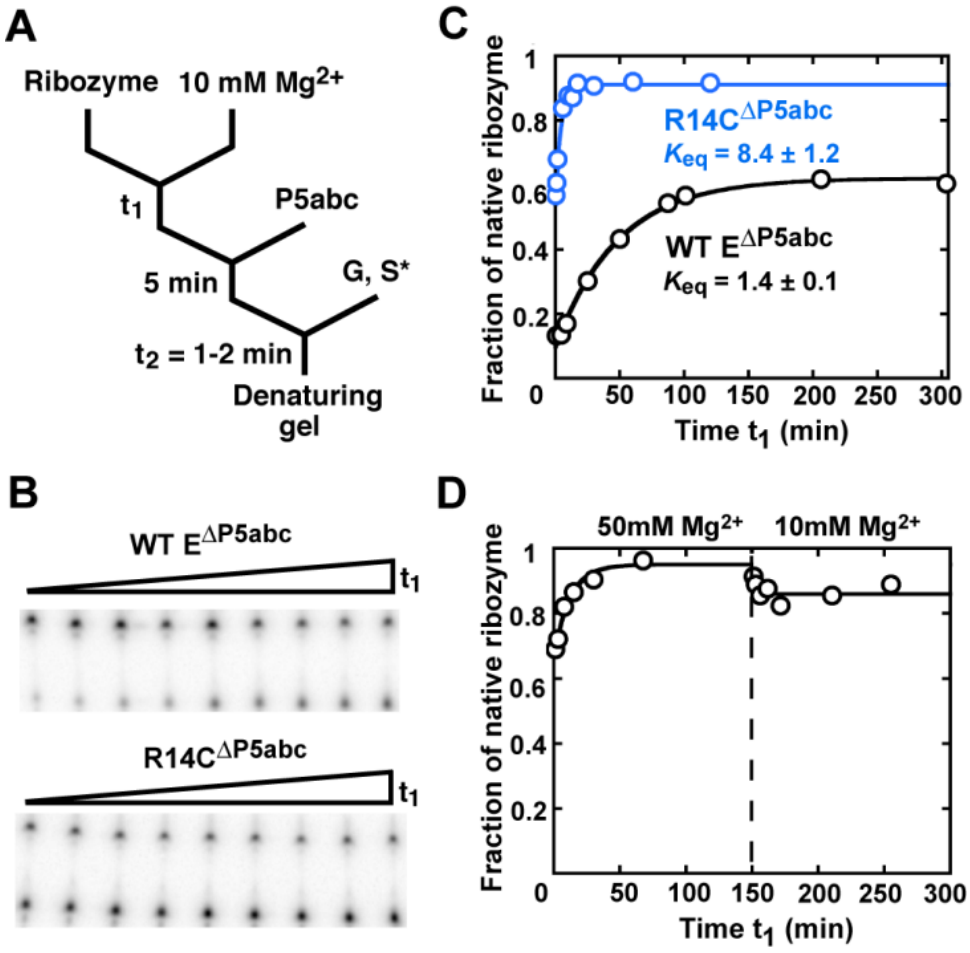

Figure 2.3: Increased specificity for native folding in the R14C$^{\Delta P5abc}$ mutant ribozyme

(A) RNA was folded in Mg^{2+} for time t_1, and then P5abc was added, followed by S* and guanosine (G). Reactions were incubated long enough to give complete S* cleavage by the native ribozyme (time t_2, 1 min for the wild-type ribozyme and 1.75 min for mutants with modestly reduced cleavage rates; see Methods). (B) Gel showing approach to equilibrium of native and misfolded conformations for the wild-type E$^{\Delta P5abc}$ (top) and R14C$^{\Delta P5abc}$ (bottom) ribozymes. Time t_1 was varied from 1 min to >250 min. (C) Folding progress curves for R14C$^{\Delta P5abc}$ (blue) and E$^{\Delta P5abc}$ (black) in side-by-side reactions. Rate constants for the approach to equilibrium were 0.25 ± 0.01 min^{-1} and 0.018 ± 0.001 min^{-1} for R14C$^{\Delta P5abc}$ and E$^{\Delta P5abc}$, respectively (average ± standard error). The latter result is the same within error as reported previously (*119*). Fractions of native ribozyme in this and other figures are calculated by dividing the fraction of S* cleaved during t_2 by the corresponding fraction for fully-native ribozyme formed by extended incubation with P5abc (>0.8, see Methods). (D) Approach to equilibrium of R14C$^{\Delta P5abc}$ upon folding at 50 mM Mg^{2+} and subsequent dilution to standard conditions (10 mM Mg^{2+}).

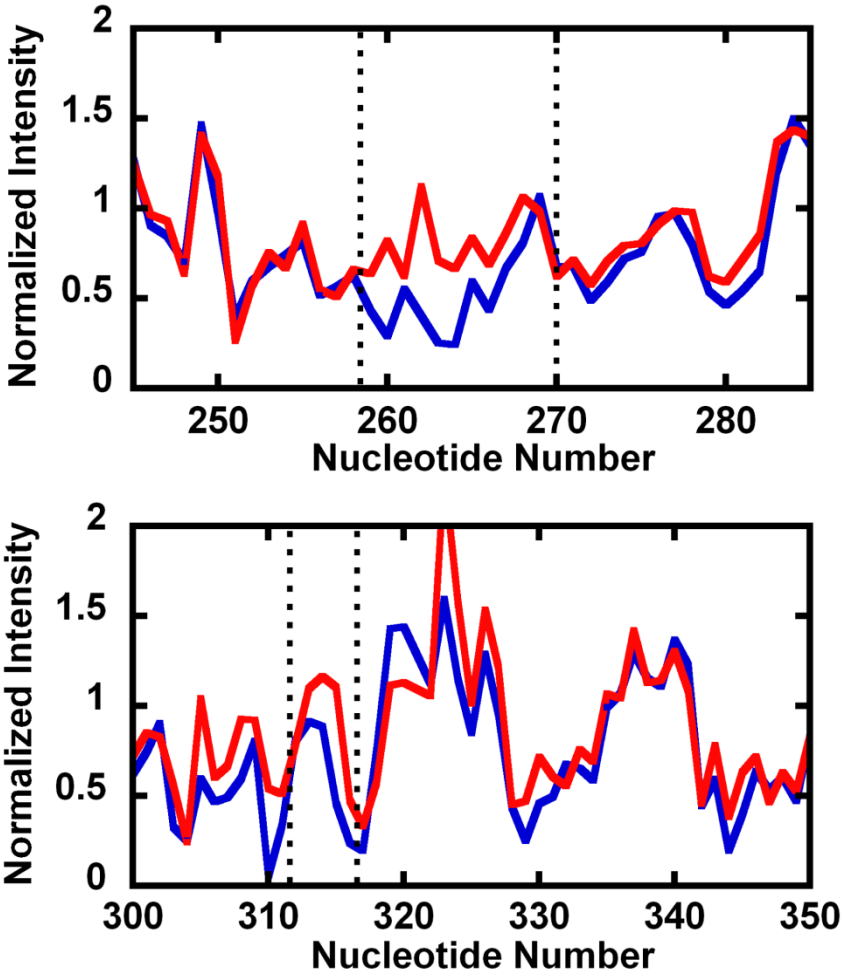

Figure 2.4: Hydroxyl radical footprinting of native and misfolded conformers of the full length R14C ribozyme

Normalized band intensities from experiments probing predominantly native (blue) or misfolded (red) ribozyme were averaged from two independent experiments. The misfolded R14C ribozyme showed features of the protection pattern that are characteristic of the long-lived misfolded species of the wild-type ribozyme (*86*). In particular, the misfolded R14C ribozyme displayed greater exposure than the native ribozyme in the regions from nucleotide 258 to 270 and from 313 to 315 (indicated by dashed lines). Footprinting experiments were performed as described previously (*86, 133*).

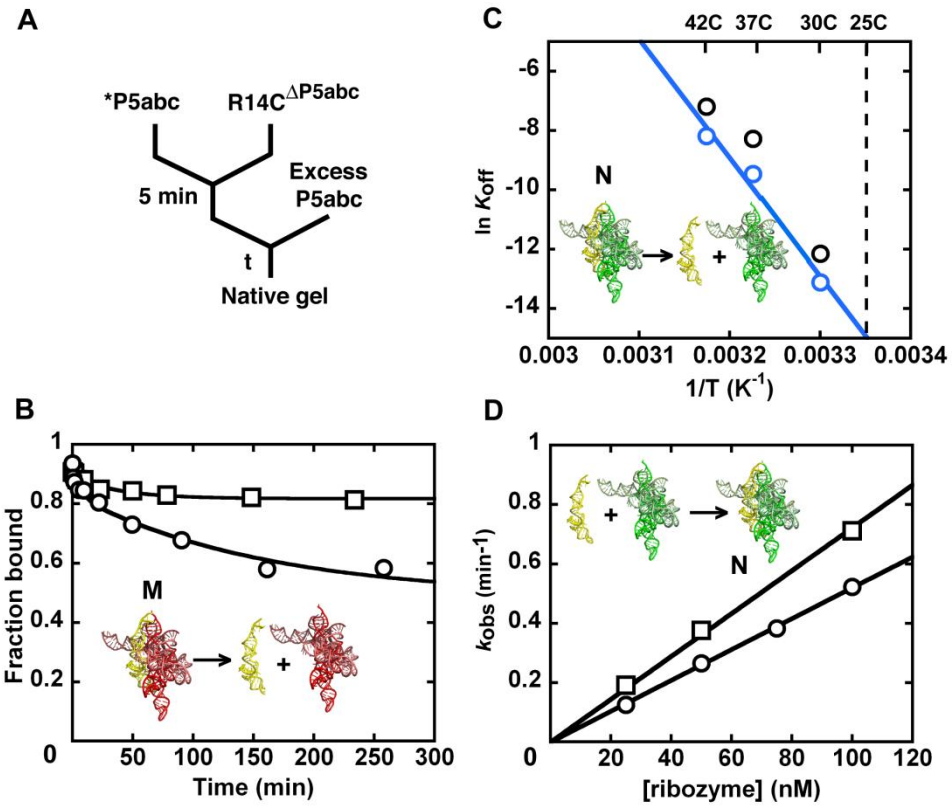

Figure 2.5: Preferential binding of P5abc to the native R14C$^{\Delta P5abc}$ ribozyme

(A) To measure dissociation, radiolabeled P5abc (*P5abc) was allowed to bind to the ribozyme and then chased with unlabeled P5abc. (B) Dissociation kinetics of P5abc from a population of predominantly native R14C$^{\Delta P5abc}$ ribozyme (squares) or a mixture of native and misfolded R14C$^{\Delta P5abc}$ (circles). The mixture gave a major phase of dissociation with a rate constant of 0.0072 ± 0.0003 min^{-1}, whereas most of the P5abc remained bound to the native ribozyme on the observable time scale (up to 5 days, data points not present). (C) Determination of the rate constant for P5abc dissociation from the native R14C$^{\Delta P5abc}$ ribozyme by extrapolation from measurements at higher temperatures (blue). Data for the E$^{\Delta P5abc}$ ribozyme obtained side-by-side are shown in black and were consistent with previous results (*119*). (D) Binding kinetics of P5abc to predominantly native R14C$^{\Delta P5abc}$ ribozyme (squares) or a mixture of native and misfolded ribozyme (circles).

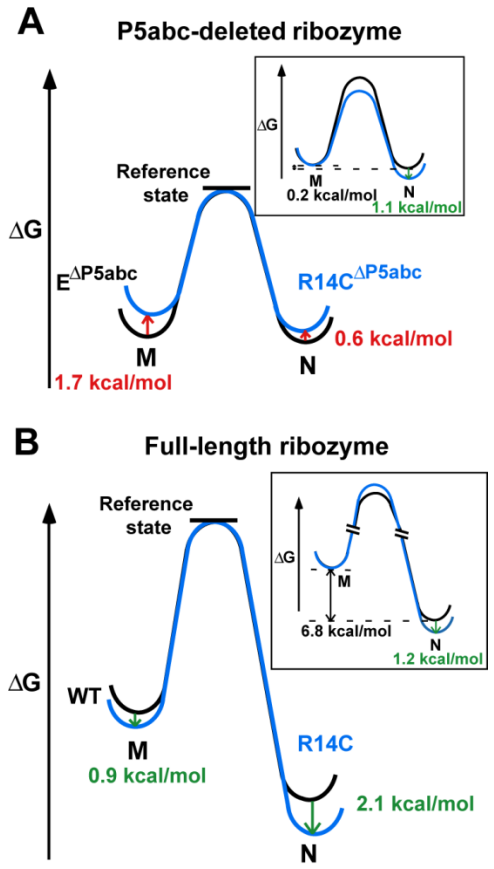

Figure 2.6: Native stabilization by the R14C mutations requires P5abc

Free energy profiles comparing wild-type and R14C ribozymes in native stability and specificity. The reference states are the relatively unstructured transition states for exchange between the native (N) and misfolded (M) conformations. For each comparison, the change in the position of M is determined by the refolding rate of the R14C variant relative to the wild-type ribozyme. The position of N is then determined from the equilibrium value between the native and misfolded conformers. (A) Folding of $R14C^{\Delta P5abc}$ (blue) and $E^{\Delta P5abc}$ (black). Both M and N are destabilized relative to this reference state (red arrows), but M is destabilized to a greater extent, giving rise to increased specificity. (B) Folding of the full-length ribozymes. M and N are stabilized by the R14C mutations (green arrows), and N is stabilized to a greater extent. In both the full length and P5abc-deleted ribozymes, the native states are stabilized by ~1 kcal/mol relative to the misfolded conformations (insets).

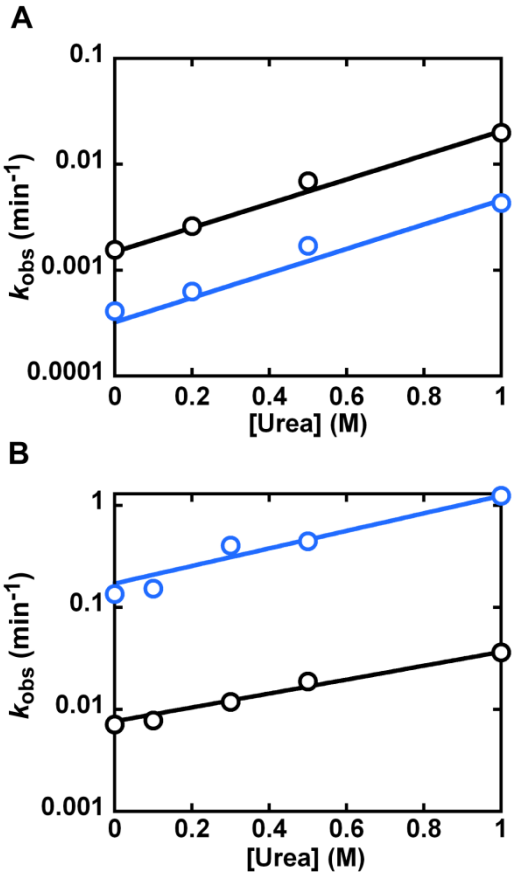

Figure 2.7: Folding transitions of the full-length and P5abc-deleted ribozymes

Urea dependence of refolding from the misfolded conformation for full-length R14C (blue) and wild-type (black) ribozymes (37°C, 10 mM Mg^{2+}, 50 mM Na-Mops, pH 7.0). The R14C ribozyme gives an m-value of 1.6 kcal mol^{-1} M^{-1}, indistinguishable from the value for the wild-type ribozyme obtained here and previously (*86*). (B) Urea dependence for refolding of the $R14C^{\Delta P5abc}$ (blue) and $E^{\Delta P5abc}$ (black) ribozymes (25°C, 50 mM Mg^{2+}, 50 mM Na-Mops, pH 7.0). The dependence for $R14C^{\Delta P5abc}$ gives an m-value of 1.2 kcal mol^{-1} M^{-1}, similar to the value of 0.93 kcal mol^{-1} M^{-1} for $E^{\Delta P5abc}$ (and the value of 1.0 kcal mol^{-1} M^{-1} measured previously, ref (*123*).

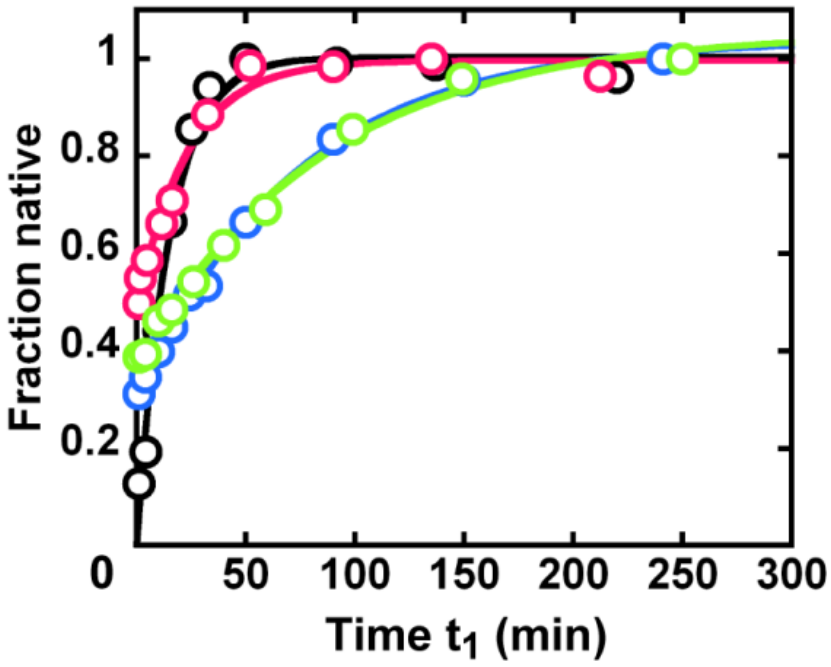

Figure 2.8: Refolding of full-length ribozyme variants

The refolding rate constant of R14C (blue) is 0.017 ± 0.004 min^{-1}, approximately 4-fold smaller than that of the wild-type ribozyme (black, 0.075 ± 0.017 min^{-1}). A variant that has only two of the nine mutations, A269G/A304G, (magenta, k_{obs} = 0.043 min^{-1}) gives a similar rate constant as the wild type ribozyme, and a variant that has all of the R14C mutations except these two (G269A/G304A, green, k_{obs} = 0.012 min^{-1}) gives a rate constant that is the same as R14C. Thus, these two substitutions do not affect the stability of the misfolded conformation relative to the transition state for refolding. Conditions were 50°C, 10 mM Mg^{2+}, 50 mM Na-Mops, pH 7.0.

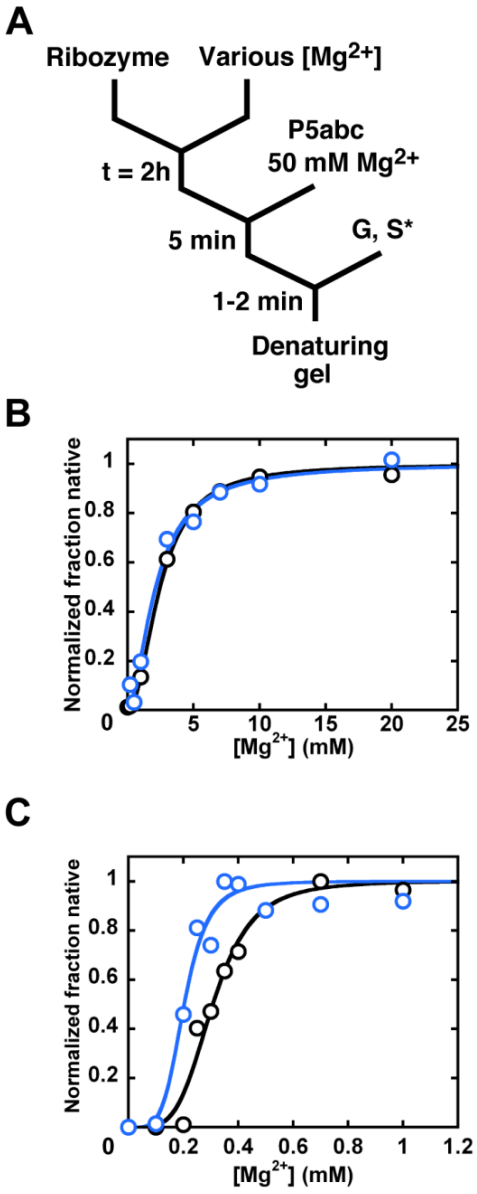

Figure 2.9: Mg^{2+} dependence of native ribozyme formation

(A) Reaction schematic for measuring the Mg^{2+} dependence (*84, 122, 123*). (B) Mg^{2+} dependences for the $R14C^{\Delta P5abc}$ (blue) and $E^{\Delta P5abc}$ (black) ribozymes are the same within error ($K_{1/2}$ = 2.2 ±0.5 and 2.3 ±0.5, respectively). To facilitate comparison, each data set is normalized to the maximum and minimum values (0.91 and 0.56 for $R14C^{\Delta P5abc}$ and $E^{\Delta P5abc}$, respectively; data not shown). (C) Mg^{2+} dependence of native state formation for the full-length R14C (blue) and wild-type (black) ribozymes. The R14C ribozyme gives a smaller $K_{1/2}$ value for native state formation (0.2 mM vs. 0.3 mM Mg^{2+}).

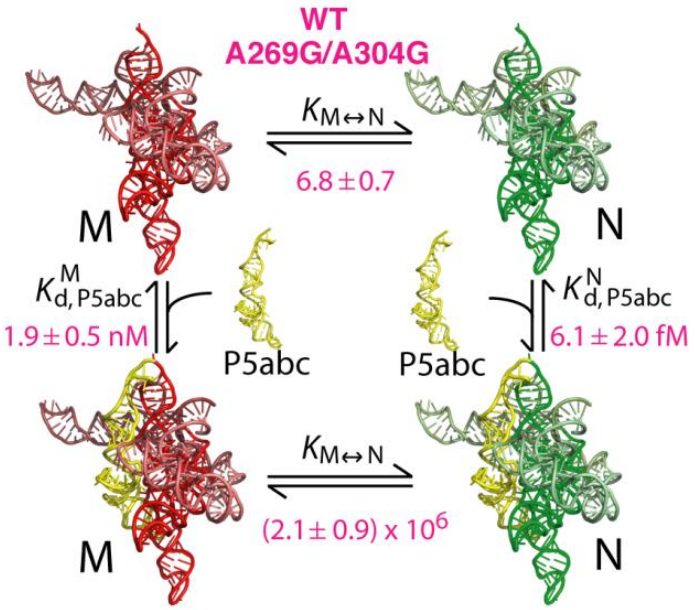

Figure 2.10: Thermodynamic cycle for the A269G/A304G ribozyme variant

Both in the absence and presence of P5abc (top and bottom, respectively), the equilibrium is shifted toward the native state by at least as large an amount as for the R14C variants (Figure 2.1, see text).

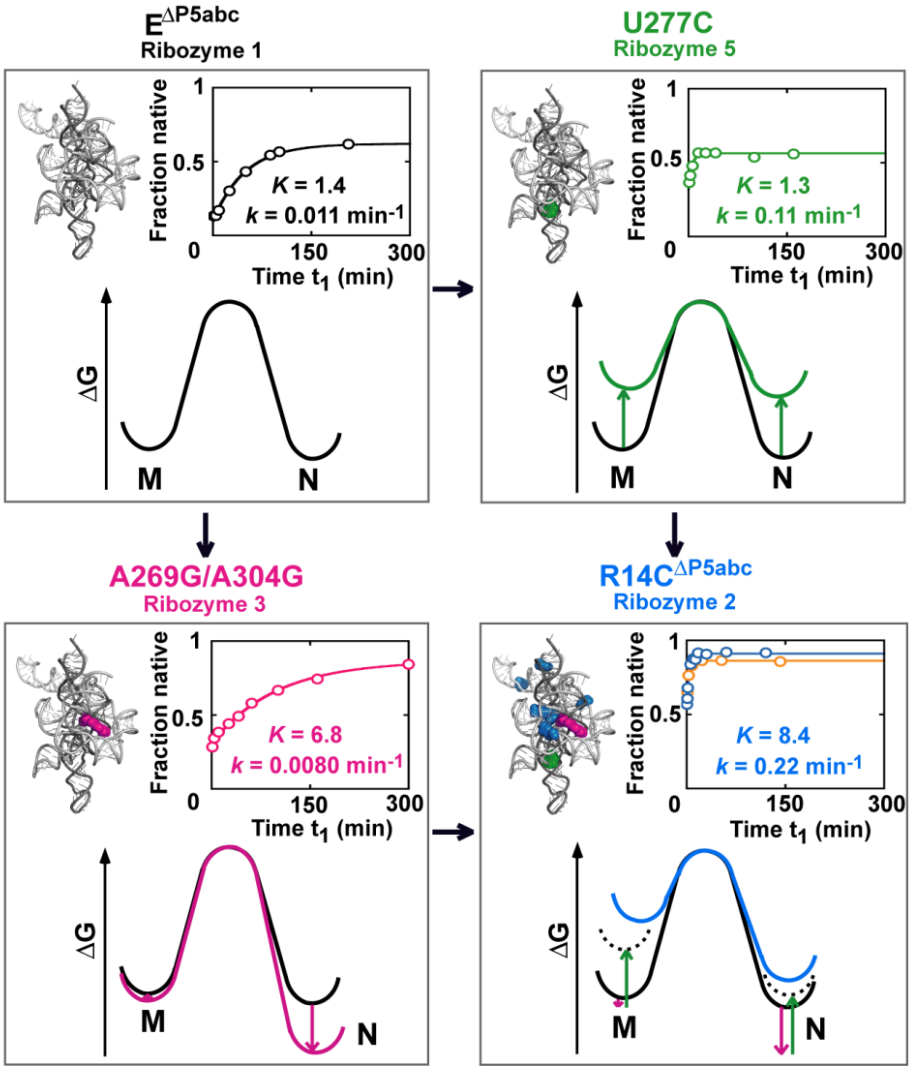

Figure 2.11: Only three substitutions affect stability and specificity in the absence of P5abc

A cycle of mutations is depicted starting from $E^{\Delta P5abc}$ (top left). Free energy profiles analogous to those in Fig 3 are shown, along with the refolding data used to calculate the profiles (insets, *k* represents $k_{M\rightarrow N}$), for the A269G/A304G variant (bottom left) and the U277C variant (top right). At the bottom right, the free energy profile and data are shown for $R14C^{\Delta P5abc}$ (blue). This ribozyme gives the same results within a factor of three from values calculated as the energetic sum of the A269G/A304G and U277C mutations, indicated by the colored arrows and dashed curves. A variant of $E^{\Delta P5abc}$ with the three mutations, A269G/A304G/U277C gives essentially the same behavior (bottom right, orange curve). Each ribozyme is labeled with its corresponding number from Table 2.1.

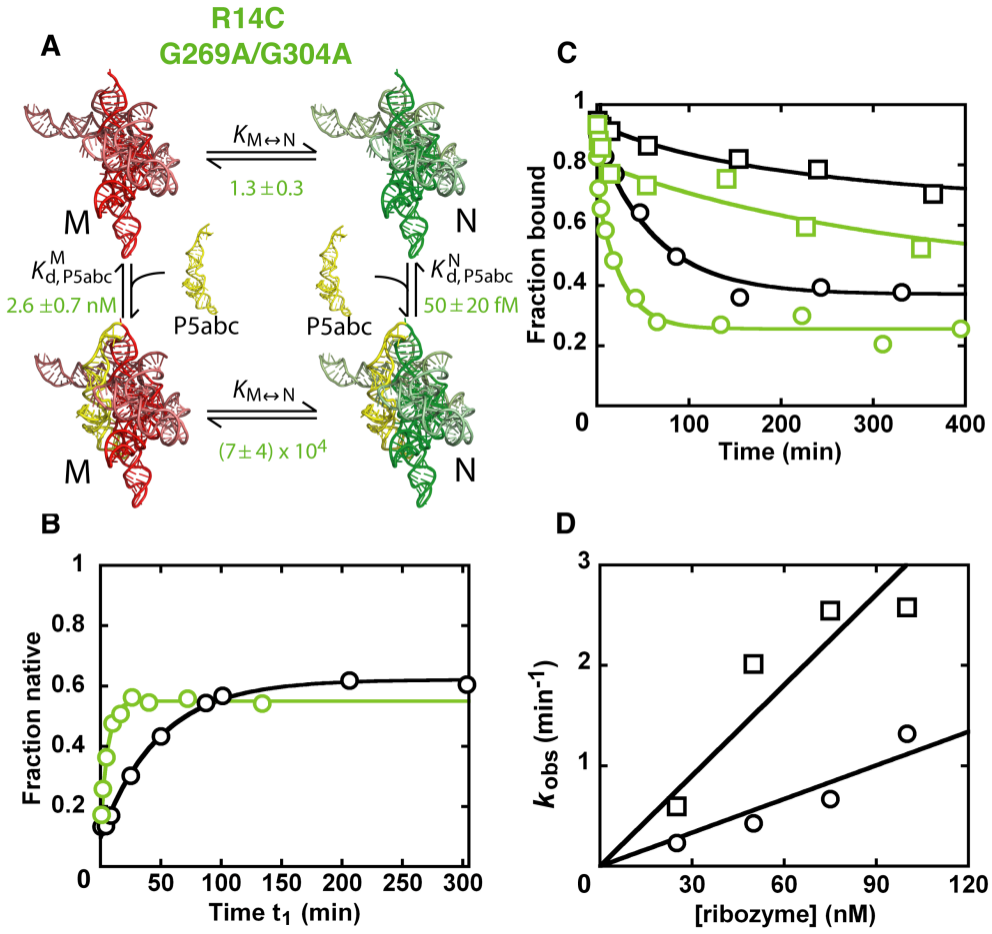

Figure 2.12: A reversion mutant G269A/G304A in the R14C background loses the enhanced specificity for native folding

(A)Thermodynamic cycle for the R14C G269A/ G304A variant. (B) Equilibrium folding of the P5abc-deleted, R14C G269A/G304A ribozyme. The observed rate constant is 0.14 ± 0.02 min^{-1} and the equilibrium value is 1.3 ± 0.3. (C) Dissociation of P5abc from P5abc-deleted R14C G269A/G304A variant (green). As in Fig S2C, dissociation was followed from a solution of predominantly native ribozyme (green squares) or a mixture of native and misfolded ribozyme (green circles). Data from wild-type ribozyme are shown in black for comparison. (D) P5abc binding curves. Analogous to experiments in Figure 2.5, binding was measured for a population of largely native ribozyme (squares) or a population of predominantly misfolded ribozyme (circles) (*119*). The calculated values from these measurements are (2.8 ± 0.6) x 10^7 M^{-1} min^{-1} for binding to the native ribozyme and (9.9 ± 1.6) x 10^6 M^{-1} min^{-1} for binding to the misfolded ribozyme.

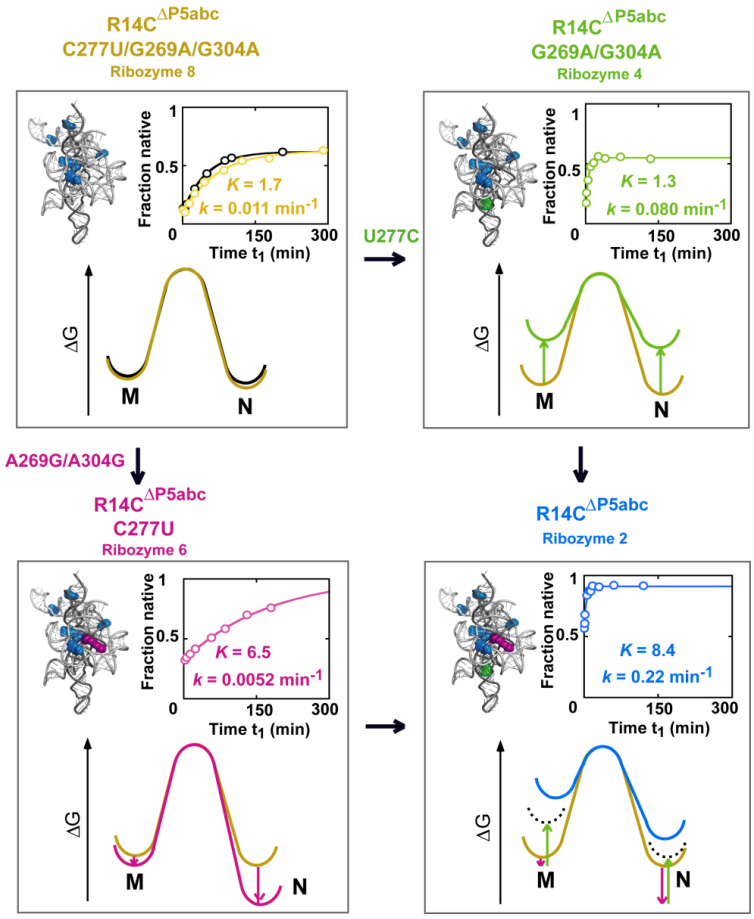

Figure 2.13: A reversion mutant G269A/G304A in the R14C background loses the enhanced specificity for native folding

Mutations of nucleotides 269/304 and 277 in the R14C background. At top left, results are shown for a mutant in which nucleotides 269,304 and 277 are reversed to their identities in the wild-type ribozyme, whereas the other six positions are mutated to correspond to the R14C ribozyme (blue in the structure schematic). Folding of this mutant is very similar to that of the wild-type ribozyme (black), as shown in the data plot (insets) and corresponding free energy profile. When nucleotides 269 and 304 are mutated to G (the 'forward' direction toward the R14C mutant), the equilibrium shifts toward the native state while the refolding rate is not affected (bottom left, magenta). The behavior of the mutant is identical to the corresponding A269G/A304G mutant in the wild-type background (see Figure 2.11). Similarly, a U277C mutant in this R14C background (top right) gives similar behavior as in the wild-type background. Predictions from the energetic sum of the effects of these mutations (bottom right, arrows and dashed curves) match the experimental results of the R14C$^{\Delta P5abc}$ ribozyme (blue) for the equilibrium value and are within 2-fold of the refolding rate.

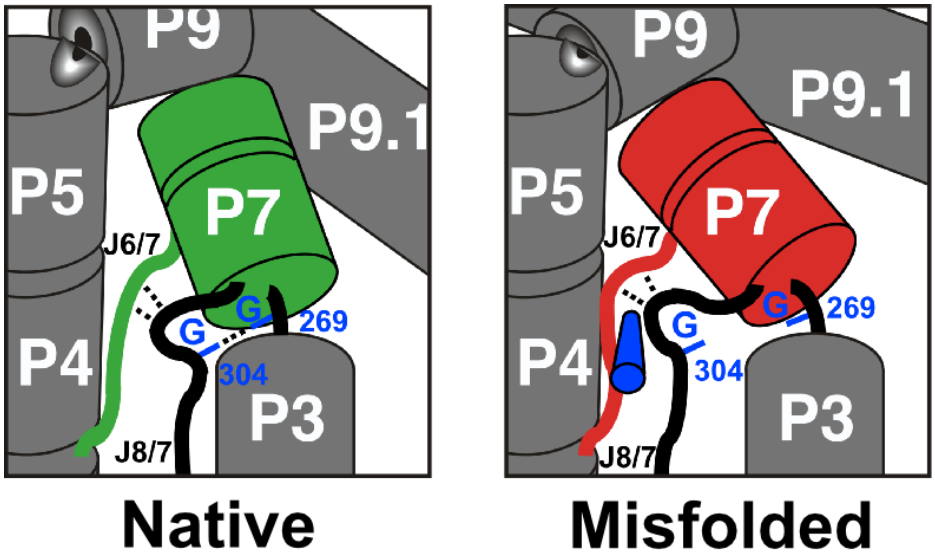

Figure 2.14: Physical model for the enhanced specificity from the A269G and A304G mutations

In the native state, guanosines at these positions can form base-specific contacts (dashed line between labeled nucleotides), analogous to contacts formed in the *Azoarcus* ribozyme between the equivalent bases (*75*, *88*). In the misfolded conformation, local rearrangements or re-orientations of P7 and J6/7 (red), as indicated by changes in footprinting patterns of these elements (*86*, *133*), change the relative positions of nucleotides 269 and 304 such that they do not contact each other. The structural differences within the core have been suggested to include a topological change (*86*), which may block contacts in the misfolded ribozyme by creating steric barriers to local rearrangements, as indicated schematically by the blue cylinder adjacent to J6/7.

	Number and description of ribozyme[2]	k_{obs} (min^{-1})[3]	$k_{M->N}$ (min^{-1})[4]	Equilibrium w/o P5abc $K_{M<-->N}$	$K_{d,}$ (M) P5abc (nM)[5]	$K_{d,}$ (N) P5abc (fM)	Equilibrium w/ P5abc $K_{M<-->N}$
1	WT[6]	0.018 ± 0.001	(1.1 ± 0.1) × 10^{-2}	1.4 ± 0.1	3.1 ± 1.0	38 ± 16	(1.1 ± 0.6) × 10^{5}
2	R14C	0.25 ± 0.01	0.22 ± 0.03	8.4 ± 1.2	2.6 ± 0.8	28 ± 10	(8 ± 4) × 10^{5}
3	WT A269G/A304G	(9.2 ± 0.6) × 10^{-3}	(8.0 ± 1.0) × 10^{-3}	6.8 ± 0.7	1.9 ± 0.5	6.1 ± 2.0	(2.1 ± 0.9) × 10^{6}
4	R14C G269A/G304A	0.14 ± 0.02	0.08 ± 0.02	1.3 ± 0.3	2.6 ± 0.7	50 ± 20	(7 ± 4) × 10^{4}
5	WT U277C	0.20 ± 0.01	0.11 ± 0.01	1.3 ± 0.1	N.D.	N.D.	N.D.
6	R14C C277U	(6.0 ± 0.3) × 10^{-3}	(5.2 ± 2.0) × 10^{-3}	6.5 ± 2.5	N.D.	N.D.	N.D.
7	WT U277C/ A269G/304G	0.70 ± 0.20	0.60 ± 0.20	6.2 ± 0.7	N.D.	N.D.	N.D.
8	R14C C277U/ G269A/G304A	0.018 ± 0.006	0.011 ± 0.003	1.7 ± 0.1	N.D.	N.D.	N.D.
9	WT A269G	0.016 ± 0.001	0.013 ± 0.001	3.6 ± 0.1	N.D.	N.D.	N.D.
10	WT A304G	(6.9 ± 0.6) × 10^{-3}	(4.9 ± 0.9) × 10^{-3}	2.4 ± 0.4	N.D.	N.D.	N.D.

Table 2.1: Kinetic and thermodynamic constants of P5abc-deleted ribozymes[1]

[1]All values are shown as the average of multiple determinations. For experiments that were performed twice, the uncertainty is expressed as the range of the two measurements. For those that were performed more than twice, the uncertainty reflects the standard error of the set of measurements.

[2]Mutants are named by first indicating either the WT $E^{\Delta P5abc}$ or $R14C^{\Delta P5abc}$ background and then indicating the nucleotide substitutions.

[3]Values are the observed rate constant for equilibration of the native and misfolded species (25 °C, 50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+}).

[4]The value of $k_{M->N}$ is calculated from the rate and equilibrium values for exchange of the native and misfolded conformations (k_{obs} and $K_{M<-->N}$).

[5]The value of $K_{d,}$ (M) and $K_{d,}$ (N) are calculated from the rate constants of P5abc dissociated and associated with the $R14C^{\Delta P5abc}$ ribozymes (data not shown).

[6]Values for the wild-type ribozyme are from side-by-side measurements except for P5abc binding rate constants, which are reproduced from previous work(*119*). All values from the side-by-side measurements were in good agreement with the earlier work(*119*).

N.D.: not determined.

	Number and description of ribozyme	k_{obs} (min^{-1})[1]			k_{obs} (relative)[2]	$\Delta\Delta G^M$ (kcal/mol)[3]	$K_{M <-->N}$[4]	$\Delta\Delta G^N$ (kcal/mol)[5]
		50 °C	37 °C	25 °C				
11	WT	0.075 ± 0.017	1.4×10^{-3}	2.1×10^{-4}	(1)	(0)	$(1.1 \pm 0.6) \times 10^5$	(0)
12	R14C	0.017 ± 0.004	4.0×10^{-3}	5.1×10^{-5}	0.22 ± 0.07	-0.95 ± 0.25	$(8 \pm 4) \times 10^5$	-2.1 ± 0.7
13	WT A269G/A304G	0.043	8.1×10^{-3}	1.3×10^{-4}	0.57	-0.36	$(2.1 \pm 0.9) \times 10^6$	-2.1 ± 0.7
14	R14C G269A/G304A	0.012	N.D.	N.D.	0.16	-1.2	$(7 \pm 4) \times 10^4$	-0.9 ± 0.9
15	WT U277C	0.041	1.5×10^{-3}	2.5×10^{-4}	0.53	-0.39	N.D.	N.D.
16	R14C C277U	8.2×10^{-3}	1.2×10^{-4}	N.D.	0.11	-1.4	N.D.	N.D.
17	WT U277C/A269G/304G	0.027	1.6×10^{-3}	N.D.	0.40	-0.59	N.D.	N.D.
18	R14C C277U/G269A/G304A	6.0×10^{-3}	1.3×10^{-4}	N.D.	0.070	-1.7	N.D.	N.D.

Table 2.2: Kinetics and thermodynamic constants of full-length mutant ribozymes

[1]Values are rate constants for refolding of the misfolded ribozyme to the native state. For the wild-type and R14C ribozymes, values are the average and range from two independent determinations. Results from single determinations are reported for the other mutant ribozymes.

[2]Relative values were obtained from measurements at 50 °C.

[3]$\Delta\Delta G^M$ values reflect the change in stability of the misfolded conformation for the indicated mutant, relative to the transition state for refolding. These values were calculated from the relative refolding rates as $\Delta\Delta G^M = -RT \ln (1/k_{obs}\text{ (relative)})$. Values are based on measurements at 50 °C. Similar effects of mutations were observed at 25 °C for all of the mutants for which measurements were made at 25 °C.

[4]$K_{M <-->N}$ values are reproduced from Table 2.1.

[5]$\Delta\Delta G^N$ values reflect the change in stability of the native conformation for the indicated mutant, relative to the transition state for refolding, and were calculated from $\Delta\Delta G^M$ and the change in $K_{M <-->N}$ as $\Delta\Delta G^N = \Delta\Delta G^M + (-RT \ln(K_{M<-->N})_{Mutant}/(K_{M<-->N})_{WT})$ (see Figure 2.6).

N.D.: not determined.

Chapter 3: Multiple unfolding events during native folding of the *Tetrahymena* group I ribozyme

3.1 INTRODUCTION

In vitro studies have shown that most RNAs are inclined to accumulate misfolded conformations that require partial unfolding to reform the functional state, typically reflected by the accelerated folding rate with the addition of denaturants (*53, 59, 80, 86, 123*). Though little detail is known about the folding process of RNA *in vivo*, misfolding and complex pathways have also been revealed (*7, 9, 67*). To understand the folding of functional RNAs thoroughly, it is crucial to be acquainted with the nature of misfolded intermediates and the transient states that lead to their formation and decay.

A model system that has been investigated extensively for RNA structure and folding is the *Tetrahymena* group I ribozyme (Figure 3.1). Williamson and colleagues used oligonucleotide hybridization to monitor the folding of this RNA *in vitro* and found a misfolded intermediate (*50*). Their later work showed that this misfolded structure is stabilized by native contacts in peripheral element P5abc (*80*). Time-resolved hydroxyl radical footprinting performed by Brenowitz and colleagues revealed more details of local structural change with tertiary structure compaction and more than one intermediate along the folding (*79, 137*). A previous study by Russell and co-workers using a catalytic activity assay to monitor folding to the native form has suggested a pathway. Upon addition of divalent ions, the unfolded RNA first compacts into an intermediate, termed

I_{trap} (*58, 83*), which is followed by another intermediate $I_{commitment}$ that diverges into a long-lived misfolded state (M) and the native state (N) (Figure 3.2).

Studies on this long-lived intermediate M tried to uncover its properties and the processes of its formation and decay. The small angle X-ray scattering experiment revealed that the global structure of M is similar to the native state (*31, 85*). Furthermore, hydroxyl radical footprinting showed that all the five native long-range tertiary contacts are formed in M and the local structure of M is identical to N except at the core (*86*). Indeed, the misfolded ribozyme can still catalyze the cleavage of the 3' mimicking substrate (*86*). However, *in vitro* refolding followed by the activity assay found that M refolds into N with a slow rate at room temperature and disruption of any of those five long-range tertiary contacts accelerates the refolding by 10-200 fold (*83, 86*). Those results naturally lead to a model where the difference between the misfolded and the native state is localized inside the core and is topological in nature, requiring extensive global unfolding for escaping from M. However, it is still unclear how and when this misfolded core forms during folding.

Therefore we concentrated on studying the I_{trap} intermediate which forms early in folding. Using hydroxyl radical footprinting as well as DMS footprinting in our lab, and activity assays performed by the Herschlag group; we discovered the structural characteristics of its local regions and the folding kinetics of the transition from I_{trap} to N. Intriguingly, we discovered that I_{trap} also forms most of the tertiary structure shared by M and N, but it has a less compact P3-P7 core than M. In addition, this early intermediate folds to N much faster than the transition from M to N and involves less structural disturbance. These results add more details to the folding pathway of the *Tetrahymena*

ribozyme and suggest multiple unfolding processes before the formation of active state of RNAs.

3.2 MATERIALS AND METHODS

3.2.1 Materials

The L-21/ScaI *Tetrahymena* ribozyme was purified using Qiagen RNeasy columns as previously described (*120*). Concentrations of wild-type and variant ribozymes were determined spectrophotometrically using an extinction coefficient of $3.9 \times 10^6\ M^{-1}\ cm^{-1}$. The oligonucleotide substrate was synthesized using standard solid-phase methods by the Protein and Nucleic Acid Facility at Stanford, 5′-end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase, and purified by non-denaturing polyacrylamide gel electrophoresis as previously described (*121*).

3.2.2 Fast Fenton footprinting

Hydroxyl radical footprinting was performed using 'Fast Fenton Footprinting' conditions (*138*). Folding of 5′- or 3′-^{32}P-labeled ribozyme was initiated at 25 °C by adding 10 mM Mg^{2+} in 50 mM Na–MOPS, pH 7.0, and 0.2% H_2O_2 and allowed to proceed for 15 or 30 s to generate a population of predominantly I_{trap} (*83, 123*); while ribozyme was also pre-incubated for 10 min to obtain 90% M or 30 min at elevated temperature (50 °C) to get mostly N. The footprinting patterns from these two incubation times were the same within error, and both sets of data were included in the reported values for I_{trap}. Footprinting was then initiated by adding Fenton reagents [5 mM $(NH_4)_2Fe(II)(SO_4)_2$ and 6.25 mM ethylene dinitrilotetraacetic acid, with Na–MOPS and

Mg^{2+} concentrations maintained at 50 and 10 mM, respectively, and H_2O_2 diluted to a final concentration of 0.15%]. After 30 s of footprinting, reactions (12 µl) were quenched by addition of 2 volume of ethanol (25 µl) and 1/10 volume of 3 M Na-acetate (1.2 µl). RNA was pelleted by centrifugation, and cleavage products were separated by 8% denaturing PAGE, imaged using a Phosphorimager (Amersham Biosciences, Piscataway, NJ), and quantified using the single-band fitting program SAFA (*139*). Intensity values were normalized relative to the average intensity from the range of nucleotides quantitated for each lane, and normalized values from three to four independent determinations were averaged (with the exception of nucleotides 30–40, which were only probed once). For each secondary-structure depiction shown in Figures 3.3-3.5, the average normalized values for each nucleotide from one ribozyme conformer were subtracted from those of another conformer to obtain difference values, which are then displayed on a color scale as indicated. In Figures 3.3D&E, nucleotide segments are highlighted in blue or yellow if at least two consecutive nucleotides had differences of at least 0.2 in their average values. From this two-nucleotide 'nucleus', the boundaries of the segment to be colored were established walking outward in each direction until either of the following were reached: (1) two consecutive nucleotides with difference values of less than 0.1 or (2) a nucleotide with a difference value of less than 0, that is, where the protection or enhancement was reversed. The DMS footprinting experiment was performed by Rick Russell and was introduced in Appendix (A.1).

3.3 RESULTS

To understand the structural properties and folding of an early kinetic trap in the *Tetrahymena* ribozyme folding pathway, I_{trap}, we performed chemical footprinting to probe the solvent accessibility in this construct side by side with the other conformations (unfolded (U), misfolded (M) and native state (N)). Further, to test the role of tertiary contacts in I_{trap} transition, the refolding kinetics from I_{trap} to N were measured for mutants and compared to the misfolded one. These experiments were performed by our collaborators in the Herschlag group, so the data is not shown in this result section but mentioned in the discussion (Section 3.4).

To probe the tertiary structure of the I_{trap} folding intermediate, we performed hydroxyl radical footprinting using Fe(II)–ethylene dinitrilotetraacetic acid (*138*). We also used dimethyl sulfate (DMS) footprinting as a probe of the base-pairing faces of adenosine and cytidine nucleotides (*140-143*). After forming I_{trap} by incubating the ribozyme briefly with Mg^{2+} ion (*83, 86, 123*), we performed footprinting for a time that is sufficiently short to prevent the ribozyme from progressing significantly to the more stable native or misfolded structures (see section 3.2.2). To allow direct comparisons between I_{trap} and other species, we footprinted the native (N), misfolded (M), and unfolded ribozyme (U) species side by side.

Changes in solvent exposure in I_{trap} relative to the unfolded ribozyme were readily apparent by hydroxyl radical footprinting (Figure 3.3A). Similar to both the native and misfolded structures, I_{trap} is protected within the core of the ribozyme (P4, P6, and P6a), as well as in peripheral elements that pack against the core (P5abc and P9.1a), and it is enhanced in exposure in several of the same regions as the native and misfolded

structures (L2.1, P5a, L6b, and P9.1). The corresponding patterns of structure formation, relative to unfolded ribozyme, are shown for M in Figure 3.3B and for N in Figure 3.4. Thus, many of the structural features of the I_{trap} intermediate appear to be similar to the native and misfolded species.

DMS footprinting also revealed substantial protection in I_{trap}, relative to U, and gave similar patterns for I_{trap} and both the native and misfolded species (Figure 3.3A&C and Figure 3.4). This similarity suggests that interconversion between these species involves at most local changes in secondary structure. DMS footprinting also provided evidence for formation in I_{trap} of at least some of the peripheral tertiary contacts. Protections were observed in loops L5b (A151–A153), L5c (C170–A172), and the A-rich bulge (A183, A184, and A186), suggesting that the three tertiary contacts involving P5abc are formed (Figure 3.1). Protection was also observed in L9.1 (A347, C350–A352), which pairs with L2.1 to form P13. Additionally, the DMS pattern of L2.1 was similar to that of the native state (*86*), with modest protection at A74 and C79 and enhanced modification adjacent to the loop (A69) relative to U. Together, these observations suggest that P13 is formed in I_{trap}.

In summary, chemical footprinting provides evidence for extensive structure formation in I_{trap}, including formation of four of the five long-range tertiary contacts. Support for the fifth contact (L9/P5) and independent support for three of the contacts (P14, L5b/J6a, and P5a/P4) comes from the mutagenesis experiments described in the next section (Section 3.4).

Despite the extensive similarities, significant differences also exist between the footprinting patterns of I_{trap} and the other folded structures (N and M). Pronounced

differences in protection from hydroxyl radicals between the native and misfolded species were reported within P7 of the core (*86*), and these differences were largely reproduced under the conditions used here (A263–A270 and U307–G312 (Figure 3.3C). In this region, I_{trap} is indistinguishable from M, and both intermediates show large differences relative to N (Figure 3.3D&E and Figure 3.5). Thus, during the transition from I_{trap} to N, the 5′-strand of P7 becomes protected from solvent, whereas in the transition from I_{trap} to M, it remains exposed. Similarly, in the localized regions that give pronounced differences between N and M in reactivity to DMS (A218–A219 in J6/6a and A342 in J9.1/9.1a), I_{trap} gives levels of exposure that are similar to M. There are also significant differences in I_{trap} relative to M that were visible in hydroxyl radical footprinting (Figure 3.3C–E). The most prominent differences are within the core, in J7/3, J3/8, J2/3, and J8/7, where each segment is more exposed in I_{trap}. There are also smaller changes in accessibility scattered throughout the structure. Thus, there is apparently a reorganization that is centered on the core during the folding transition from I_{trap}. Consistent with this model, these regions were shown in time-resolved footprinting experiments to be protected on the time scale that would be expected for further folding from I_{trap} (*79*). DMS footprinting revealed additional significant differences in modification of a small number of nucleotides (Figure 3.3C and Figure 3.5A). Nucleotide C255, near the interface of the P4–P6 and P3–P8 domains, was protected in M and N relative to I_{trap}, raising the possibility of a rearrangement at or near the domain interface. Nucleotides C278 and A103, within P3, and A306, within J8/7, were also protected from DMS modification in M and N relative to I_{trap}, providing further support for a rearrangement in the P3–P8 domain of the core.

In summary, the observed changes in solvent exposure of core elements suggest that there is substantial reorganization of the core subsequent to I_{trap} formation. The high degree of similarity of the DMS footprinting patterns of I_{trap}, M, and N suggests that there are no large differences in secondary structure. Indeed, even the 'unfolded' state ('U' in Figure 3.2) is thought to include most of the secondary-structure elements present in the final native RNA structure (*131*).

3.4 DISCUSSION

More properties of this early kinetic trap have been uncovered by monitoring its transition to N with denaturant and mutants with disrupted long-range tertiary contacts. These experiments were done by a graduate student, Hyejean Suh, in the Herschlag lab. A brief description of the assay and results is given in the Appendix (A.2). Briefly, the results showed that urea modestly accelerates escape from I_{trap} to N, indicating much less structure disruption in this transition than the transition from M to N. Meanwhile, disruption of any of the five long-range tertiary contacts accelerates transition from I_{trap} to N with an effect less than 10- fold; whereas it was accelerated by 100- to 1000- fold in transition from M to N. Thus, all of the data suggested that I_{trap} folds into the native structure without extensive unfolding. In addition, mutants in L2.1 and L9.1 that breaks P13 interaction outside the P3-P7 core gave no effect for I_{trap}, indicating that unfolding events might focus on the P4-P6 side of the ribozyme during the transition from I_{trap} to N.

3.4.1 Escape from I_{trap} via local unfolding

Differences of solvent protection pattern between intermediates and the native state, as well as the effects of denaturant and disruption of tertiary contacts, assist the refinement of the *Tetrahymena* ribozyme folding pathway. Escaping from I_{trap} was found to disrupt only 5-6 base pairs, and the disturbances are localized mainly in the P4-P6 domain which participates in four long-range tertiary contacts. On the other side of the ribozyme, disruption of the P13 tertiary contact, formed between L2.1 and L9.1 around the core region (P3-P7), gave no obvious acceleration of refolding kinetics. Interestingly, hydroxyl radical footprinting revealed that the main differences between I_{trap}, M and N are positioned within the core, near the P13. Therefore, a simple model to reconciling these results is that unfolding events occur at the P4-P6 region during folding from I_{trap}, but they help the compaction of the overall molecule, especially the core area on the other side (Figure 3.6).

The P4-P6 domain was demonstrated previously to form early in the folding pathway, even prior to I_{trap} (*50, 79, 144, 145*). Earlier studies showed that folding of the P4-P6 domain involves close packing of the P5abc subdomain against the P4-P5-P6 subdomain, and folding can occur independently of the rest of the ribozyme (*77, 146*). The isolated P4-P6 domain is stabilized by two tertiary contacts, L5b-J6a/b and P5a-J4/5, which were shown to form cooperatively with an equilibrium constant of 12 at 22 °C, 10 mM Mg^{2+}, and 200 mM Na^{+} or 100 at higher Mg^{2+} (70 mM) (*132*). Assuming that the stabilities of these tertiary contacts are the same during folding of the entire molecule, one would imagine that the complete disruption of these contacts should introduce the same scale of acceleration in the folding kinetics. But we only observed around a 5-fold

increase in the refolding from I_{trap} to N in the mutants (50 mM Mg^{2+} and 25°C). Indeed, urea dependence showed the involvement of only 5-6 base pairs during the transition from I_{trap} to N. Together, these data suggested that breaking down the tertiary contacts in P4-P6 facilitates the transition from I_{trap} but is not a necessary event. It is very likely that there are multiple pathways to escape from I_{trap} because disruption of each of four tertiary contacts all speed up folding but only few base pairs are involved in this step. Such scenario of complex folding landscape has also been observed earlier (*54, 131, 132, 147*).

3.4.2 Multiple unfolding steps in continued folding from intermediates to the native structure

As folding proceeds, I_{trap} forms $I_{commitment}$, which then partitions into the native and misfolded forms. I_{trap} is a loose compact kinetic trap that forms early in the pathway. It provides more space for the RNA to break nonnative contacts and search for native contacts. A small fraction of I_{trap} forms the native conformation while the rest is trapped in a more compact kinetic trap, M. Although M is structurally similar to the native form except for a topological barrier within the core, it slowly refolds to the native state with extensive unfolding that breaks nearly all of long-range tertiary contacts (*86*).

We noticed that structures of I_{trap} and M are also similar to each other; however, their folding properties are distinct. Footprinting data showed that the main differences between I_{trap} and M are located in the core. The ribozyme core is protected in M relative to I_{trap}, and is further sequestered from the solvent during the transition from M to N, suggesting increased packing in the molecule's center. In addition, formation of the five peripheral contacts impedes the folding of I_{trap} modestly while preventing the folding of

M extensively (*86*). These results suggest a model in which the peripheral contacts are strengthened as folding progresses from I_{trap} to M, probably arising partially from increased cooperativity. An example for long-range cooperativity between tertiary contacts came from the study of the peripheral element P5abc in our lab (reference (*119*); T. Johnson and R.R., unpublished data). Tertiary contacts formed by this P5abc element and other tertiary contacts in the ribozyme generate a large cooperativity that may contribute to the enhanced structural stabilization of P5abc to N relative to M (6 kcal mol^{-1}).

All of the results suggest that the cooperativity within the RNA increases with its folding. The formation of interconnected and cooperative tertiary contacts may cause additional challenges in RNA folding, requiring RNAs to take multiple rounds of folding and unfolding as they traverses rugged energy landscapes. Thus, the complexity of energy landscapes for RNA folding increases with the complexity of the RNA structures. Local rearrangement and multiple rounds of unfolding may govern folding of large structured RNAs.

This chapter was adapted from published paper:

Wan, Y., Suh, H., Russell, R., and Herschlag, D. (2010) Multiple unfolding events during native folding of the Tetrahymena group I ribozyme, *J Mol Biol 400*, 1067-1077.

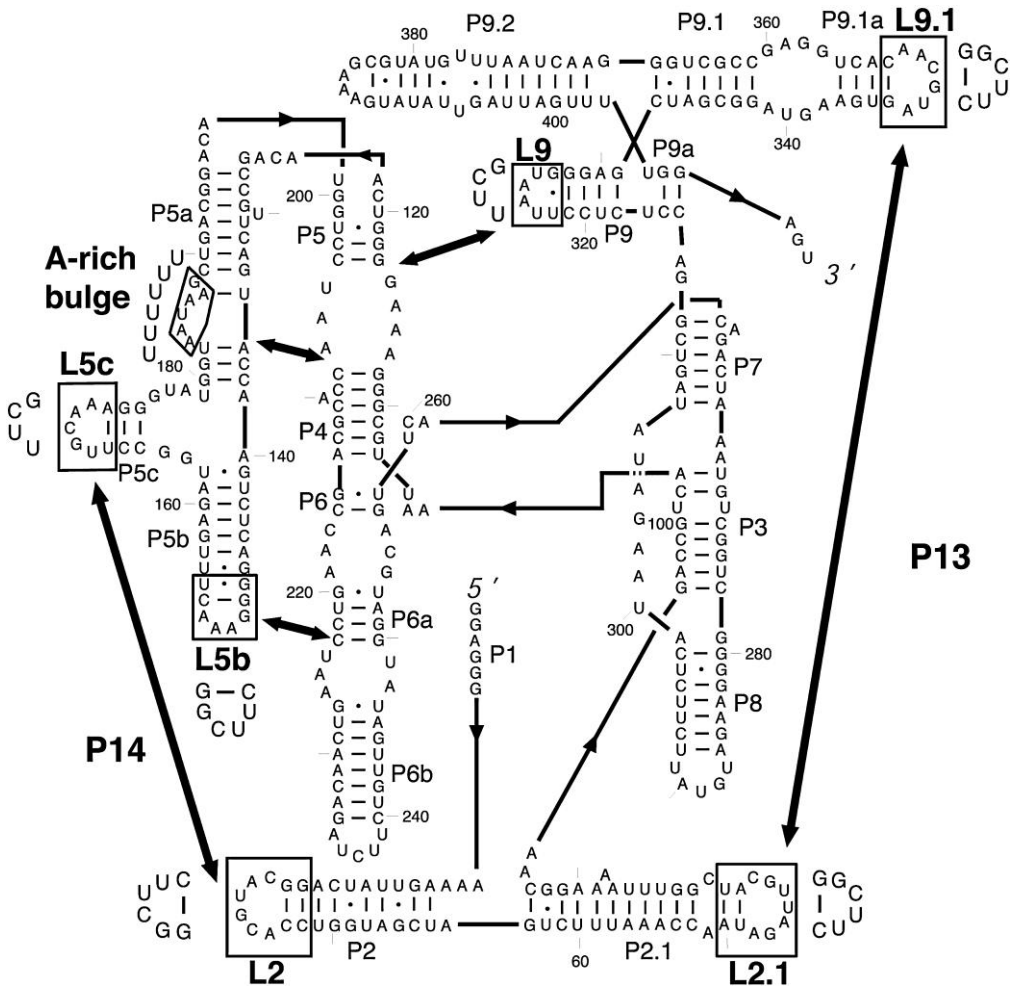

Figure 3.1: Secondary structure of the *Tetrahymena* group I ribozyme

The five long-range contacts are indicated by thick arrows, and the mutations made to abolish these tertiary contacts are indicated with boxes, with the substituted residues shown adjacent to each box.

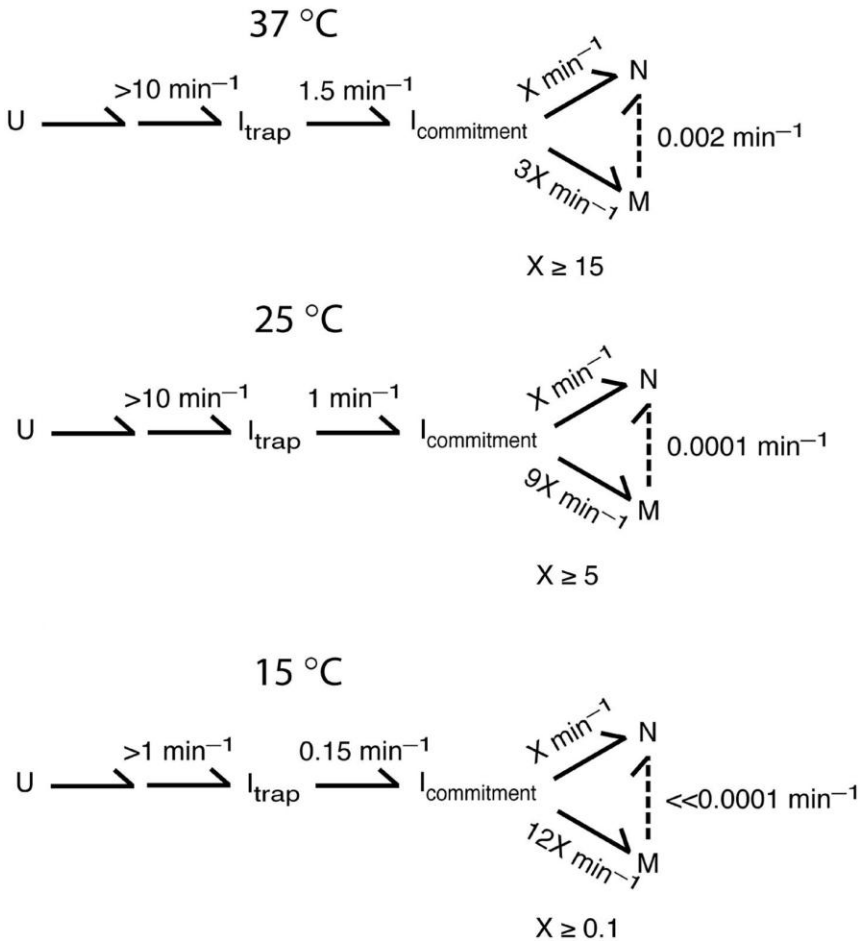

Figure 3.2: Folding schemes for the *Tetrahymena* ribozyme under different temperatures

Rate constants for each folding step of the *Tetrahymena* ribozyme are showed under 37 ℃, 25 ℃ and 15 ℃ for comparison (*83*).

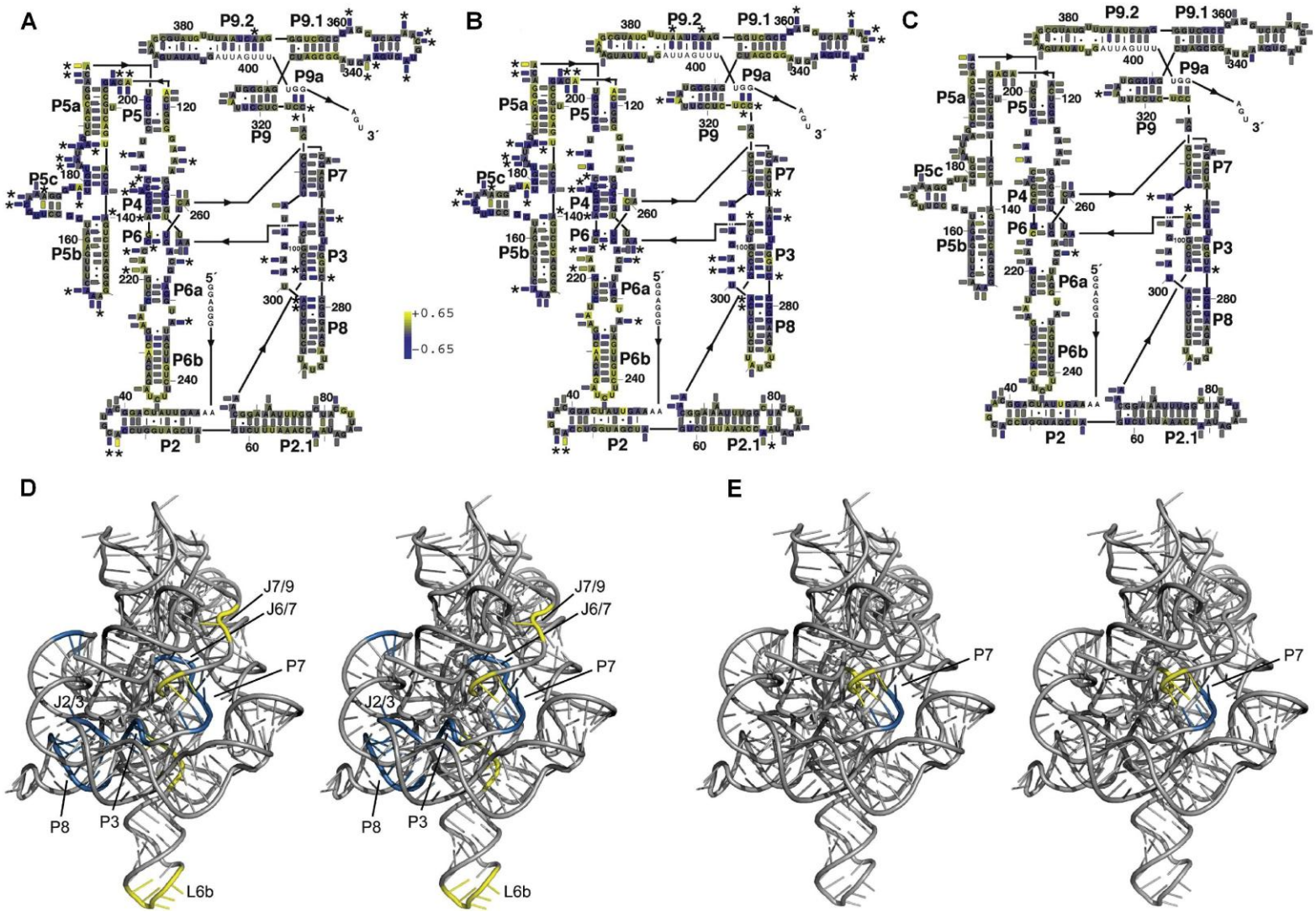
A
B
C
D
E
P9.2
P9.1
P9a
P5a
P5
P5c
P4
P6
P5b
P6a
P6b
P7
P3
P8
P9
P2
P2.1
+0.65
-0.65
J7/9
J6/7
J2/3
L6b

Figure 3.3: Hydroxyl radical and DMS footprinting of the I_{trap} folding intermediate

(A–C) Comparisons of footprinting pattern for different intermediates displayed on the secondary structure. Shaded boxes beneath the nucleotide letters represent results of hydroxyl radical footprinting, and boxes next to the letters represent results of DMS footprinting (*86*). The color of each box corresponds to the difference in band intensity and, therefore, the difference in exposure to the chemical probe of the indicated nucleotide between the two ribozyme species in the comparison (see Section 3.2.2). (A) Comparison of I_{trap} with the unfolded ribozyme. Nucleotides that are protected in I_{trap}, that is, modified by the probe to a lesser extent, are blue, and those that are exposed in I_{trap} are yellow. (B) Comparison of M and U. Nucleotides that are protected in M are blue and those that are more exposed in M are yellow. (C) Direct comparison of M and I_{trap}. Nucleotides that are protected in M relative to I_{trap} are blue, and those that are exposed in M are yellow. In (A)–(C), the nucleotides with the most prominent changes in DMS reactivity are highlighted with asterisks (Appendix A.1). (D and E) Comparisons of hydroxyl radical footprinting of the native ribozyme with I_{trap} (D) or M (E), with results displayed on the model of the native tertiary structure. Only the most prominent changes are displayed, with regions that are protected in the native state shown in cyan and those that are enhanced shown in yellow (see Section 3.2.2).

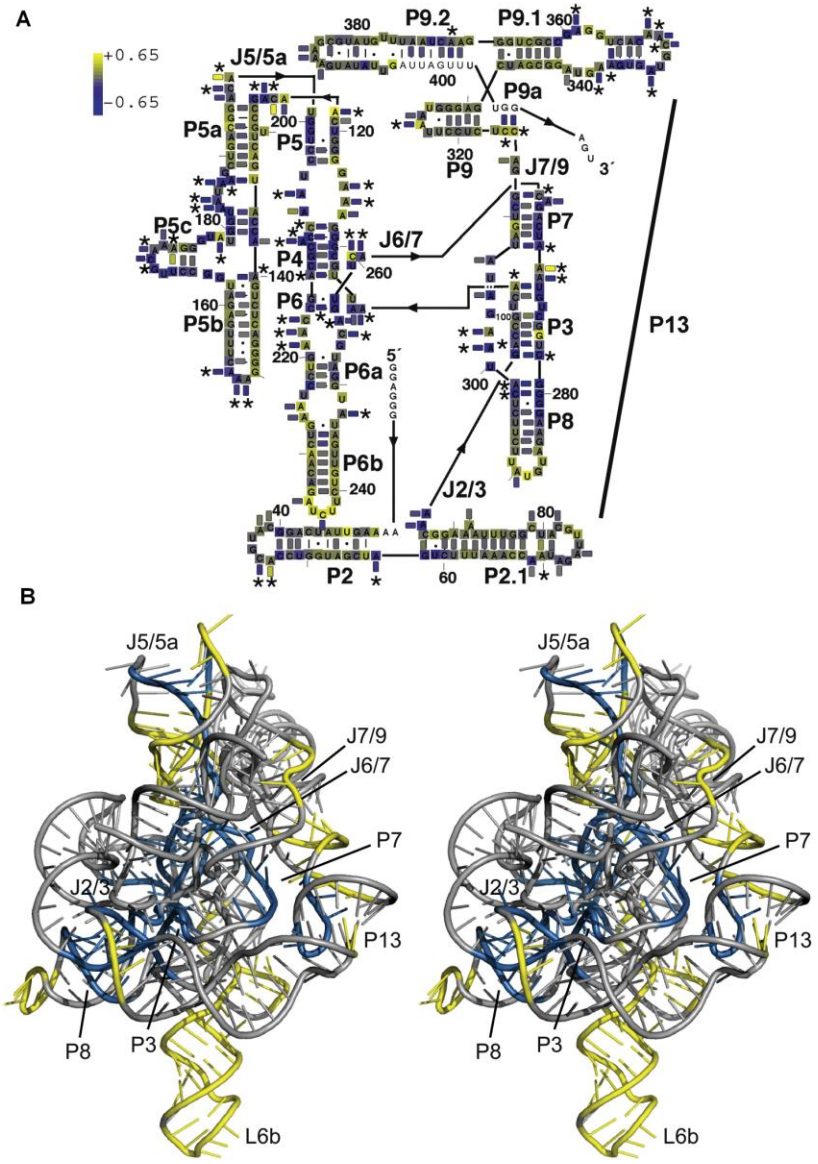

Figure 3.4: Hydroxyl radical and DMS footprinting comparisons of the native state and unfolded state

(A) Comparisons of footprinting pattern for different intermediates displayed on the secondary structure. Positions that are more protected in the native species are shown in blue, and those that are more exposed in the native species are shown in yellow. (B) Stereo view of the three-dimensional model with the most prominent protections from hydroxyl radicals in N shown in blue and the most prominent enhancements shown in yellow.

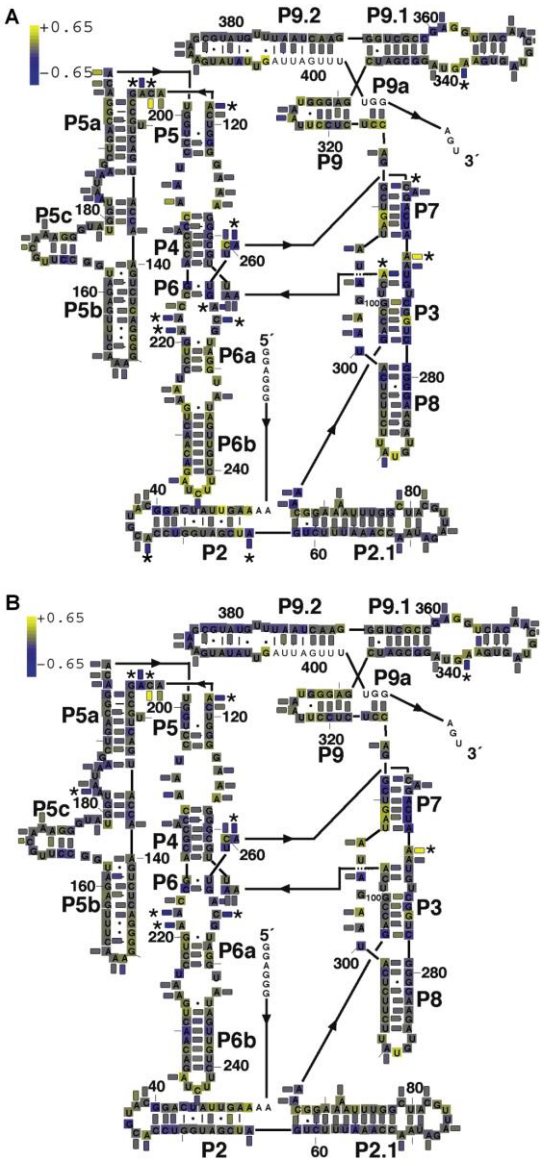

Figure 3.5: Comparisons of footprinting pattern for the native ribozyme and earlier intermediates displayed on the secondary structure

(A) Comparison of native ribozyme and the I_{trap} intermediate, with results shown on the secondary structure. (B) Comparison of the native and misfolded ribozyme (M). For each comparison, positions that are more protected in the native species are shown in blue, and those that are more exposed in the native species are shown in yellow. In all panels, the nucleotides with the most prominent changes in DMS reactivity are highlighted with asterisks.

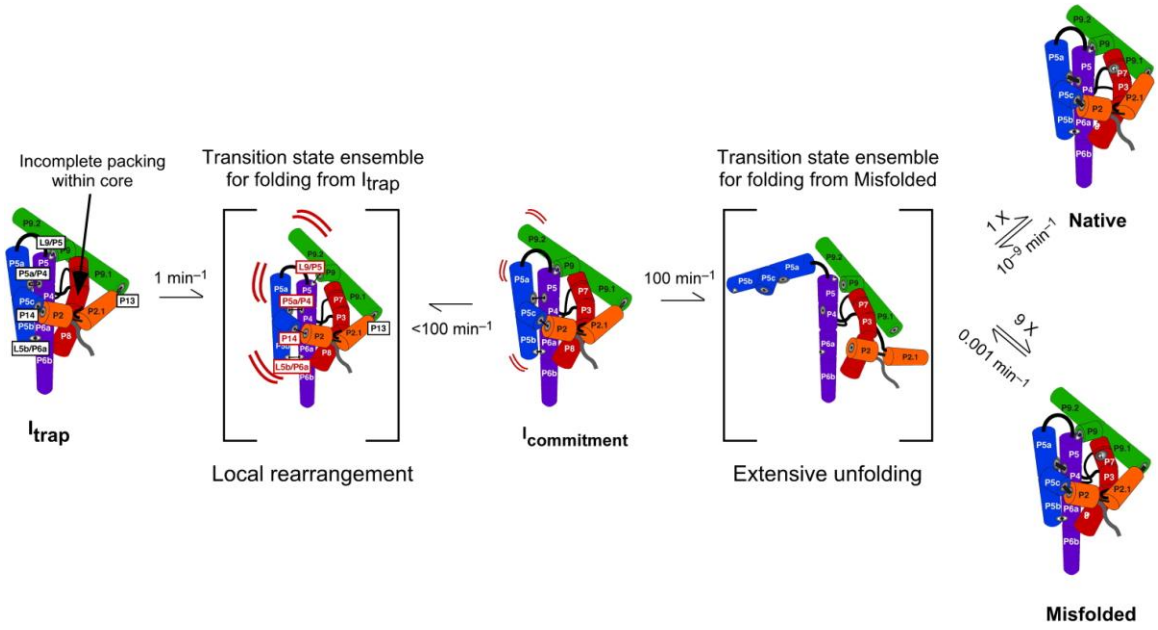

Figure 3.6: Model for continued folding of the intermediates I_{trap} and M

The five long-range tertiary contacts are labeled in boxes. The four contacts that accelerate folding from I_{trap} when mutated are labeled in red in the transition state adjacent to I_{trap}. P13 apparently remains formed in the transition state from I_{trap} to the subsequent intermediate $I_{commitment}$ (indicated by the black label). $I_{commitment}$ then continues to unfold more globally and partitions predominantly to M, which then slowly re-folds to N. Folding transitions between $I_{commitment}$, M, and N likely proceed through a large collection of intermediates and transition states that are unfolded to varying degrees and may or may not be the same for the different transitions. For simplicity, the entire collection is shown as a single transition state. The rate constant of 100 min^{-1} from $I_{commitment}$ to this transition state was determined previously by measuring folding starting from conditions that allow avoidance of I_{trap} but not $I_{commitment}$ (*31*). M then slowly unfolds back to this transition state with a rate constant of 0.001 min^{-1}, giving an observed rate constant of 10^{-4} min^{-1} after accounting for the 10-fold greater likelihood of returning to M rather than proceeding to N. Once reached, the native ribozyme unfolds back to this transition state very slowly. The indicated rate constant of 10^{-9} min^{-1} is calculated from the measured value of 10^5 for the equilibrium between N and M (*119*). All of the rate constants shown are from experiments at 25 °C and 10 mM Mg^{2+}.

Chapter 4: Extensive Tertiary Contact Formation and Novel Architecture in a Group I RNA

4.1 INTRODUCTION

Since the discovery of an RNA intron molecule that possessed catalytic activity in 1982 (*70, 148*), interest has been put on the investigation of functional RNAs. They have been found to play important roles in tRNA and mRNA maturation, protein translocation, as well as mRNA translation with the help of proteins *in vivo* (*2, 3, 149*). To possess their functions, many RNA molecules, such as RNase P RNA, ribosomal RNA, self-splicing introns and sets of small RNAs have to fold into a specific three-dimensional structures (*20, 150, 151*). RNA structure became an important field to explore in the recent decades.

Large structured RNAs such as RNase P or ribosomal RNA share a similar strategy in their three-dimensional structure that can be exemplified by the group I self-splicing intron, in which a conserved core is generally enclosed by sets of peripheral helices that form long-range tertiary contacts to each other. The group I intron is a good model system to study RNA structure because over 2000 sequences are known and they share a similar construction strategy. Among 5 classes and 13 subclasses of group I RNAs, they all share a conserved core that is composed of two helical domains, P4-P6 and P3-P7, and have versatile peripheral helices buttressing this conserved core (*72, 73, 76*). Only a few of them have been studied extensively and differences between them has already provided us with profound insights to the architecture of RNAs (*75, 77, 78, 88, 89, 152*). For instance, in the *Tetrahymena* group IC1 ribozyme, there is a P5abc helix that forms

three tertiary interactions with its adjacent helix P4-P6 and another peripheral element P2; meanwhile, in the *Azoarcus* group IC3 ribozyme, P5abc is missing, and P2 instead stacks on P8. These differences reflect how RNAs seek balance in growing structural complexity. However, it is still unclear why those elements are kept and how they are designed by nature to stabilize the functional structure. Thus, exploring the diversity of natural architecture of those elements will broaden our knowledge of RNA structure and functions, as well as helping to predict the tertiary structures of RNA.

Nevertheless, the exploration of other group I introns is restricted to developments in X-ray crystallography because RNA molecules have dynamic architectures and form numerous intermediates. To date there are about five group I introns that have been reported by themselves or with proteins in high-resolution crystal structures (*75, 78, 88, 89, 93*). Scientists thus need to seek other biochemical or biophysical means to resolve the 3-D structure directly or indirectly. The benefit of nuclear magnetic resonance (NMR) and small X-ray angle scattering (SAXS) is that both of them can detect the structure of RNA molecules in solution. However, NMR is deficient in solving structures for large RNA molecules, while SAXS falls short at presenting macromolecule structure at atomic level. Therefore, diverse biochemical methods have been employed to determine RNA structures indirectly. Solvent exposure assays such as oligonucleotides hybridization, footprinting experiments (hydroxyl, DMS, etc.) are powerful in probing the local structure of folded molecule. For those structured RNAs that are able to catalyze a reaction, their structures can be quantitatively monitored by catalytic assays. Additionally, computational modeling programs have been growing enormously to bridge

the gap between geometric growths of known functional RNAs and limited available structures.

In this paper, we are using biochemical methods and computational modeling to study a new group I intron, located at the position 516 of the *Bangia* (red algae) nuclear-encoded small subunit (SSU) 18S rRNA. It belongs to the IC1 group of introns, the same category as the *Tetrahymena* group I RNA. Comparative analysis predicted that all *Bangia* group I introns have the typical primary and secondary structure features of the group IC1 subgroup (*102, 103*). However, there are two novel structural features uniquely belonging to this group I intron: a three-base pair P8 helix capped by GNRA tetraloop, and an extra stem near P5b that is 6 base pair long and capped by GNRA tetraloop (termed as P5b.1) (Figure 4.1A). They are pervasive and conserved in most of *Bangia* group I introns, but not present in other group IC1 introns (*102, 103*). Previous evidence has shown that most GNRA tetraloops that are capped with a fixed length helix would form tertiary contacts (J. Lee and R. Gutell, unpublished data), thus it is very likely that one or both of these novel elements contribute to tertiary contacts.

To establish the structure of a group IC1 ribozyme from *Australia Bangia*, we used biochemical methods including mutagenesis and catalytic activity assays as well as hydroxyl radical footprinting. The results suggest that the ribozyme possesses most of the tertiary contacts formed in the *Tetrahymena* ribozyme. Interestingly, we discovered that the *Bangia* ribozyme forms a contact between L8 and P2, similar to the contact found in other subgroups in group I intron RNAs (e.g. group IC3 and IA2), but the tertiary elements have been swapped. In addition, this contact was confirmed by a three-dimensional model that was generated from the biochemical data and homologous

structural segments from other group I introns by a computer program. Thus, our study has extended the concepts of modularity of RNA tertiary motifs, as well as provided a general strategy to explore the unknown tertiary structures of RNAs.

4.2 MATERIALS AND METHODS

4.2.1 Preparation of the wild-type and mutant *Bangia* ribozymes

The *Bangia* intron DNA was cloned from strains obtained from K. Muller (University of Waterloo, Ontario, Canada). RNA was first tested to be active in self-splicing reactions and then constructed into ribozyme versions by removing both 5' and 3' exons. The mutant ribozymes were obtained by Quikchange mutagenesis PCR. Wild type and mutant ribozyme RNAs were prepared by *in vitro* T7 polymerase transcription and purified as described (*58, 120, 153*). The RNA oligonucleotides were synthesized by Dharmacon Research, Lafayette, CO and were radio-labeled at the 5'end with [γ-^{32}P] ATP. The *Bangia* ribozyme was 5′-end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and 3′-end-labeled with [α-^{32}P]ATP using Klenow fragment, as described in Section 2.2.6 (*121, 127*).

4.2.2 Cleavage activity assays

The ribozyme (final concentration 400 nM) was preincubated at 50 mM Na-MOPS, 10mM Mg^{2+}, pH7.0, 500 μM guanosine at 50°C for 30 minutes to form the native structure, then 5' end radiolabeled substrate (UCUGGUGAAC) was added, aliquots were taken out at various times and quenched by 70% formamide, 100mM EDTA, 0.1% bromophenol blue, and 0.1% xylene cyanol. Labeled substrate and product (UCUGGU)

were separated by running on 20% denaturing PAGE and quantitated by a Phosphorimager (Molecular Dynamics).

4.2.3 Static small angle X-ray scattering

The experiment was performed at the beam line 12-ID of the Advanced Photon Source (APS) at Argonne National Laboratory in Chicago. The energy of X-ray was selected at 12 keV throughout all whole measurements. All the assays were performed at 25°C using a static cuvette (volume = 20 µl). Ribozymes (2 µM) were pre-incubated at pH 7.0, 50mM Na-MOPS, varying Mg^{2+} (0-10 mM), 50°C for 30 minutes before injecting into the cuvette (The free Mg^{2+} concentration is lower than the number presented here because the RNA molecule takes up Mg^{2+} ions). Buffer blanks were also performed in parallel. Each sample or buffer blank was measured by five sequential exposures of 0.2 s. These data were then averaged and the scattering from the buffer blank was subtracted from the RNA scattering curve.

4.2.4 3D modeling of the *Bangia* ribozyme

The RNABuilder program developed by Flores Samuel and Russ Altman in Stanford University and available for downloading from simtk.org was used to generate models of the *Bangia* ribozyme tertiary structure (*154*). The modeling used the *Tetrahymena* ribozyme structural model generated by Lehnert and Westhof and part of the crystal structure of the *Azoarcus* ribozyme from Adam et.al. (1ZZN) as a template (*76, 88*). We first put in the primary sequence and the start residue number. Then we modeled the residues of a helix in the Bangia ribozyme onto the corresponding residues in the

template by command 'superimpose'. The secondary structure information predicted from comparative analysis was put in after the threading step for each helix using command 'WatsonCrick'. Regions that do not have a template we modeled only by the latter command. In addition, spheres were added to residues in single-stranded regions to avoid strand cross. Tertiary contacts that have been proven by the mutagenesis and activity assay were entered in the end, including L2.1-L9.1, L9-P5, P5bulge-J4/5 (see text). Drawings of the obtained 3D models were produced by Visual Molecular Dynamics (VMD) (*155*).

4.2.5 Simulating ideal solvent accessibility data and SAXS data

The simulated solvent accessibility profile of each C4' atom on the backbone of both RNA models was generated by the software Surface Racer (*156*), with a radius of 1.6 Å for the solvent probe. The output data was normalized and compared to experimental generated hydroxyl radical footprinting data. The simulated SAXS data for each model was obtained by running them on the program CRYSOL (*157*).

4.3 RESULTS

4.3.1 The *Bangia* ribozyme possesses most of the characteristic tertiary motifs of group I introns

We started to explore the global structure of the *Bangia* group I intron by specifying its tertiary elements. To study the structure of this novel group I RNA using a catalytic activity assay as a read-out, we first designed the ribozyme version of the *Australia Bangia* group I intron (Figure 4.1), which only catalyzes a *trans* substrate mimicking the

first step of splicing. The substrate of this ribozyme contains a sequence that is complementary to the IGS (internal guide sequence) and the 5' splicing site. *Trans* cleavage assays were conducted under standard conditions (25°C, pH7, 10mM Mg^{2+}, 50mM Na-MOPS, 500uM guanosine) after pre-incubating the ribozyme at 50°C for 30 minutes. The cleavage rate of the Bangia ribozyme is 0.016 min^{-1} (Figure 4.2), which is much lower than that of *Tetrahymena* ribozyme but still robust enough to be monitored.

To ensure that the cleavage step is rate limiting in the single turnover assay, we varied the reaction conditions including concentration of the cofactor guanosine, pH, concentration of Mg^{2+}, and the pre-incubation time with Mg^{2+}. The guanosine concentration dependence curve gave a $K_{1/2}$ of 180 μM (Figure 4.3A), which is similar to the K_D of the *Tetrahymena* ribozyme and also confirmed that guanosine was saturated at standard conditions (500 μM). We also found that the cleavage rate increased log-linearly with the pH (pH 6.0-pH 8.0), indicating that the chemical step is rate-limiting under standard conditions (pH 7.0) (Figure 4.3B). We varied the concentration of Mg^{2+}, the cleavage rate constant increased steeply at low concentrations (0.5-5 mM Mg^{2+}), but linearly at higher concentrations (10-100mM Mg^{2+}) (Figure 4.3C). Finally, in order to make sure that the folding of native ribozyme has achieved equilibrium before cleavage assay, we changed the pre-incubation at 50°C. The result not only showed that 15-20 minutes was sufficient to allow the approach to folding equilibrium (data not shown) but also indicated the presence of a misfolded species in the *Bangia* ribozyme.

Even though the *Bangia* ribozyme has a similar secondary structure as the *Tetrahymena* ribozyme, its cleavage activity is much lower. One possibility is that the P1 duplex is not docking properly in the active site. We tested this by measuring the binding

affinity of the substrate with saturating cofactor guanosine (Table 4.1). The K_D of substrate to the ribozyme is equal or smaller than 3 nM under 25°C, whereas K_D of substrate to the IGS alone calculated by Nearest-neighbor rule (*158*) is 4 nM. Since the real K_D is systematically smaller than the calculated one (*159*), substrate binding seems to not be enhanced much by presence of the ribozyme. This similarity indicated that P1 docking is inefficient in the *Bangia* ribozyme, leading to the slow cleavage rate in the single turnover assay.

In addition to activity assays, small angle X-ray scattering (SAXS) was performed to ensure that the ribozyme folds into a compact structure (Figure 4.4). Radius of gyration (R_g) of the *Bangia* ribozyme decreased upon addition of magnesium ions, from 71 Å in the absence of Mg^{2+} to 44 Å with 1.5 mM Mg^{2+}, similar to the previous data from the *Tetrahymena* ribozyme (*85*). The R_g value increases by a small amount from 1.5 mM to 10 mM Mg^{2+} as observed before for the *Tetrahymena* ribozyme (*85*), probably reflecting a small amount aggregation.

The *Bangia* ribozyme has most of the typical tertiary elements found in group IC1 RNAs (*102, 103*). To test whether these elements actually form tertiary contacts in this novel group I ribozyme, we constructed several mutants that disable the potential contact individually and tested the catalytic activity of these mutants. Generally, we mutated the candidate loops by replacing the sequence with a UUCG tetraloop, which still forms a stable loop structure but does not recognize the corresponding receptor of the original tetraloop. The only exception is the A-rich bulge in the P5a stem, which was mutated to a poly U sequence. All the activity tests of mutants were performed under standard conditions as described previously for the wild-type *Bangia* ribozyme.

Measurement of the mutant activity revealed that most of the tested tertiary motifs are essential for full-activity of the *Bangia* ribozyme (Figure 4.5, Table 4.2). Mutations of loop L2.1 and L9.1 both decreased the cleavage rate by 10-30 fold, suggesting that these loops form the long-range interaction P13 like the other group IC1 ribozymes, although the mutation in L2.1 gave a more severe effect. This is probably because L2.1 is near the active site and its modification would cause more problems than breaking a long-range tertiary contact. Mutations in L9 and the P5 A-rich bulge, on the other hand, reduce the activity of the wild-type ribozyme by 5-10 fold, suggesting that both of them form interactions with their partners. That is, while L9 interacts with P5, the P5 bulge docks on J4/5.

Interestingly, we noticed some differences between the *Bangia* ribozyme and the *Tetrahymena* ribozyme. For instance, the L5b tetraloop interacts with J6a/b in the *Tetrahymena* ribozyme and its mutant decreases the activity of ribozyme (*81*), however, in the *Bangia* ribozyme, the L5c mutant has the same activity as the wild-type ribozyme. To confirm this result, we brought down the Mg^{2+} concentration so the ribozyme was less stable so that the subtle influence arising from the mutation could be observed. However, even at 0.5-1 mM Mg^{2+}, the L5b mutant still showed the same cleavage activity as the wild type ribozyme, indicating that this mutation did not affect the ribozyme structure much or this L5b element is not involved in any tertiary contacts. One possibility is that the novel L5b.1 loop takes the role of L5b loop, which is discussed in the following section.

Another intriguing observation came from the mutant of the L5c loop, which exhibited a 10-fold reduced activity, indicating that it forms a tertiary interaction and

contributes to the structure stability; on the other hand, the other half of this interaction is not straightforward to identify because there is no complementary sequence in the L2 loop as observed in the P14 contact of the *Tetrahymena* ribozyme. Additionally, the sequence of the L5 loop is UAAC, which is not a typical tetraloop that has been normally studied and thus the knowledge of its potential receptor sequence is inadequate.

To gain a more specific scope of the local structure of folded RNAs, we used hydroxyl radical footprinting as a complementary tool. In the footprinting assay, RNA is initially folded in metal ions to adopt a specific conformation; chemical reagent is added subsequently to generate hydroxyl radicals, which cleaves the exposed C4 atoms on the backbone. The resulting RNA fragments are separated on a denaturing PAGE gel and the relative band intensity indicates solvent accessibility for individual residues. Hydroxyl radical footprinting reports the protection pattern of each residue on the tertiary structure level, but not secondary base-pairing information. To probe structure of the *Bangia* ribozyme, we first folded the RNA at 10 mM Mg^{2+}, 50 mM Na-MOPS, pH 7.0 and 50 °C for 30 minutes to approach the equilibrium of native ribozyme formation. Then we added Fe(II)-EDTA footprinting reagents at 25°C to generate hydroxyl radicals. After 10 minutes cleavage, the reaction was quenched by thiourea and formamide (see Section 2.2.6) and run on an 8% PAGE gel. For comparison, unfolded RNA in the buffer that lacks Mg^{2+} ions was probed in parallel. When comparing the protection data of the folded ribozyme with the unfolded control, we observed protections at the majority of the earlier determined tertiary elements, including L2.1, L5c, P5 budge, L9, L9.1, as well as the ribozyme core P3-P7 (Figure 4.6). Noticeably, even the L5b element, whose mutant has

the same activity as the wild type was protected. It is likely that the loop of L5b is formed and becomes sequestered during global folding.

4.3.2 Evolution kept the tertiary interaction but swapped the elements

After probing the overall structure of the RNA, we set out to test the novel tertiary elements in the *Bangia* ribozyme that do not exist in other group IC1 introns. Again, we disrupted the putative contacts by mutating the two novel tetraloops, L8 GCAA and L5b.1 GCAA into UUCG tetraloop, and tested the catalytic activity of these two mutants. Additional mutations were made to change the tetraloops into GCGA tetraloops, which may interrupt the potential tertiary contact but with moderate effect. It was shown that the L8 UUCG and GCGA mutants both dramatically decreased cleavage rate constant (e.g. 2.9×10^{-4} min^{-1} for the UUCG mutant), compared to the wild-type ribozyme (0.014 min^{-1}) (Figure 4.7). To rule out the possibility that the reduced activity at high Mg^{2+} was due to weak binding of substrate or guanosine, we varied the ribozyme and guanosine concentration for the L8 mutant and found both were saturating at standard conditions (data not shown). In contrast, the L5b.1 UUCG mutant had nearly the same activity as the wild-type ribozyme at standard conditions (Figure 4.7). But it showed less activity at lower Mg^{2+} concentrations, suggesting that L5b.1 might form a contact in tertiary structure, which could be compensated by other tertiary contacts if deleted, similar to P13 and P14 mutations in the *Tetrahymena* ribozyme (*82*). Put together, these results gave insights into the critical role of the L8 tetraloop in the global structure of the *Bangia* ribozyme and thus led us to ask which part was contacted by L8.

Helices that are close to P8 were all taken into consideration, and P2 and P2.1 were the strongest candidates. Alignment of P2 and P2.1 from more than 60 related *Bangia* group I intron sequences revealed two consecutive GC base pairs near the base of each stem (Figure 4.1B). Interestingly, those two consecutive GC base pairs are conserved in RNAs which also contain tetraloops L5b.1 and L8, but are not conserved in others. Studies of receptors with GNRA tetraloops have also shown that consecutive GC base pairs are potential receptors for GYAA tetraloops (*160-162*). Thus these regions in P2 and P2.1 are substantial candidates for the receptor of the novel L8 tetraloop. Additional support came from the associated group IC3 and group IA2 intron structures, although the elements are switched in these RNAs and a L2 GNRA tetraloop docks on the P8 helix.

In order to identify the receptor, we mutated both consecutive GC base pairs in P2 and P2.1 to UA:UA so that the potential tertiary contact would be disrupted. The reason we chose this mutation sequence is it changes the local geometry of the base-pairs to the largest extent, while keeping the helix structure at the same time. Remarkably, the result showed that mutating P2 in the wild-type background slowed down the cleavage rate by about 100- fold, which is similar to the behavior of the L8 UUCG single mutant (Figure 4.8). A mutation that replaced only one of the consecutive GC base-pairs in P2 has also been tested and showed no obvious decrease in activity (data not shown), indicating tolerance of the tetraloop. On the other side, P2.1 mutant gave a similar rate as the wild type ribozyme. It is apparent that the P2 receptor site has more potential to form tertiary contacts than the P2.1 site. To further probe the potential L8-P2 contact, we constructed a double mutant containing both mutation sequences. We found that mutating P2 in the L8

UUCG background did not decrease the cleavage rate further, suggesting that P2 and L8 form a tertiary interaction with each other so that the double mutant has the same activity as a single mutant.

To rule out the possibility that the cleavage activity was too low to decrease further in the double mutant and the likelihood that what we obtained was the background activity of the ribozyme we did several sets of controls. We first tested the activity of another double mutant containing the P2 UA:UA and the L5c UUCG tetraloop. It was found that this double mutant further decreased the activity compared with the single mutant under the same conditions, proving that the rate constant of the L8/P2 double mutant did not reflect the background activity (Table 4.3). In addition, it also indicated that L5c does not interact with this putative P2 receptor site. The second control experiment measured the activity of all the double-mutant and single-mutant ribozymes at a higher Mg^{2+} concentration (100mM Mg^{2+}), in which the difference between mutants could be enlarged and readily observed. As we expected, the results confirmed that the L8-P2 double mutant has the same activity as single mutants (Table 4.3).

A useful strategy to confirm the existence of a tetraloop-receptor contact is to restore the contact via a compensating double mutant. In principle, replacing the wild-type tetraloop with an alternative GNRA loop may reduce the cleavage rate moderately; this decreased activity can be restored by changing the receptor sequence specifically for this new GNRA loop. The limitation of this strategy is that the specific receptor sequence for each GNRA tetraloop has not been understood thoroughly. It renders the design of the double mutant extremely difficult.

We first designed a combination of L8 GCGA and P2 CU : AG in the mutant, which is a typical tetraloop-receptor interaction (*163*). L8 GCGA mutation decreased the cleavage rate, but not as severely as the L8 UUCG mutation (Figure 4.7); meanwhile, P2 CU:AG presented the same activity as the wild type ribozyme, suggesting that this sequence is tolerable for the wild-type L8 loop (GCAA). The double mutant of this tetraloop-receptor did not restore the activity of ribozyme to the level of wild type (data not shown), indicating that a typical tetraloop-receptor contact might not form under different context geometry. We further designed several sets of compensation mutants according to work from Costa and Michel (*160*) (Figure 4.9). They *in vitro* selected receptors of different GNRA tetraloop, and the system they used was the L2 GNRA tetraloop and the P8 receptor in the *td* group IA intron. None of our mutants showed significant restored activity, even under varying conditions (Mg^{2+} concentrations and temperature) (data not shown). However, this did not argue against our conclusion since the tetraloop-receptor interaction has not been studied extensively. A design that tried to replace a known tetraloop-receptor contact (L9-P5) in the *Azoarcus* ribozyme also failed to restore the ribozyme activity (*44*). Moreover, the reactive helix P1 stacks on the P2 duplex and a slight change in the duplex length of P2 and its position may diminish the ribozyme activity to a large extent and interfere with the data interpretation.

4.3.3 Three-dimensional model of the *Bangia* ribozyme

To reconcile all biochemical data and confirm the novel tertiary contact between L8 and P2, we generated a 3D model for the *Bangia* ribozyme by a computational modeling program RNABuilder developed by Russ and Flores (*154*). RNABuilder uses the internal

coordinate dynamics code together with a largely coarse-grained potential and fine-grained kinematics (*164*). Instead of utilizing the Cartesian space of individual atoms, it takes advantage of internal coordinates, which propagate the dynamics of a chain of bodies by varying the parameters of its connection points, e.g. bond length between atoms and the torsion angles. Therefore bonds between certain atoms can be fixed (those in the same base planar) in each nucleotide, while the remaining rotable ones are calculated. The benefit is that RNABuilder can enforce any input base-pair interactions characterized by Westhof group in 2002 (*165*). Analogously, base stacking interaction and coaxial stacking can be built into the model easily. Another feature that derived from this principle is that a known piece of RNA structure from crystallization or modeling can be readily rigidified as a whole unit without losing the molecular details inside it. Sequence of new homologous RNA can be then superimposed (or 'thread') onto conserved structures with little computational cost. All these advantages give us an opportunity to model large RNA molecules that are longer than 300 nucleotides, which would otherwise take a long time in Molecular Dynamics modeling.

In a previous paper, we applied RNABuilder to model the *Azoarcus* group I intron ribozyme (*122*). The strategy was to use only the secondary structure information of *Azoarcus* ribozyme and thread it onto a template mainly composed of the crystal structure of the *Twort* ribozyme by matching the corresponding helices of the two ribozymes. The L9-P5 and L2-P8 tetraloop-receptors in this *Twort* structure have been replaced by the GAAA tetraloop-11nt receptor from the *Tetrahymena* ribozyme crystal structure. The model we obtained by RNABuilder has a RMSD of 4.6 Å compared to the crystal structure of the *Azoarcus* ribozyme, indicating the accuracy of the RNABuilder software

in modeling large RNA structure. The success of predicting a large group I intron RNA by the threading strategy allows us to model more unknown structures for other RNAs, especially the group I introns.

We applied a similar approach to generate a 3D model of the *Bangia* ribozyme on the base of homologous group I RNAs. We used the structure model of the *Tetrahymena* ribozyme (it has been validated by crystal structures of the major domains and biochemical data (*75-78*)) as the template since the *Bangia* ribozyme highly resembles the *Tetrahymena* ribozyme as shown by biochemical results. But there are two stems (P5b.1 and P8) in the *Bangia* ribozyme that are distinct from the template RNA as shown in the biochemical data. Since the L5b.1 mutant reduces ribozyme activity modestly, whereas the L5b mutant has no effect, we decided to 'thread' this functional P5b.1 to the P5b stem in the template and left the three-base-pair P5b stem sticking out from the main coil.

On the other hand, the positions of the P2 and P8 stems are more complicated. Data from the double mutant L8-P2 has suggested a tertiary contact that is not observed in the *Tetrahymena* ribozyme, in which the P2 stem turns to the P5abc domain and forms the L2-L5 tertiary interaction. Earlier work on other introns has revealed a similar contact in the *Azoarcus* ribozyme and some other group I ribozymes (*Twort*, bI3), in which P2 coaxially stacks on the P1 duplex, and the end of the P2 helix (L2) docks onto the P8 duplex by tetraloop-receptor interaction. However, whether P2 stacks on P1 as that in the *Azoarcus* ribozyme or on P2.1 like in the *Tetrahymena* ribozyme is not clear. First, the tertiary motifs have been swapped in the *Bangia* ribozyme relative to those in the *Azoarcus*. Second, RNA tertiary motifs are modular and independent of context, and the

similar connection between two elements does not necessarily require the adjacent helices to be oriented in the same way.

Therefore, we attempted modeling with both scaffolds; one composed of the entire *Tetrahymena* RNA structural model and the other possessing most of the *Tetrahymena* architecture but with P1/P2 from the *Azoarcus* ribozyme (Figure 4.10). Since the P8 helix of the *Bangia* ribozyme contains only three stems and is highly restricted to the covalently linked P3 stem, we used the first three stems of P8 in the *Tetrahymena* ribozyme as the template in both scaffolds, but no force has been applied to connect L8 and the putative receptors. Additional challenges came from structural elements that are not present in any solved structures, such as P5d region and the L8 loop; those regions can be then modeled by RNABuilder.

For modeling, we also put in structural information including secondary structure and tertiary contacts (L9-P5, L2.1-L9.1, L5b.1-J6a/6b, and P5a-J4/5) that have been demonstrated from biochemical experiments. Two models of the *Bangia* ribozyme were generated after several rounds of running. In the first model, P2 was threaded onto the same helix in the *Tetrahymena* ribozyme and coaxially stacks on P2.1. The other end of P2 turns to the L5c loop without contacting because there is no P13 contact between L5c and L2 in the *Bangia* ribozyme (Figure 4.11). Remarkably, the putative receptor in P2, two consecutive GC base pairs sit in the minor groove of the helix in this model and are readily to form tertiary contacts. The short P8 stem in this ribozyme ends right next to this P2 receptor site. The GCAA tetraloop flips upwards so that it sits within the range of forming interactions with the P2 receptor.

However, this contact is not favored in the second model. We found that when P2 coaxially stacks on P1, the receptor GC base pairs are hidden in the shallow groove and far from the P8 helix (Figure 4.11). Actually, the lower end of P2 is closer to L8 and suggests an alternative candidate receptor there. To further test this possibility, we shortened the P2 stem to only four base pairs and capped it by a UUCG tetraloop. The alternative site is apparently deleted from this mutant. Activity assay showed that this mutant has the same cleavage rate constant as the wild type ribozyme (Figure 4.12), strongly disfavoring the idea of the P2 end being involved in any tertiary contacts. In addition, the L5c mutant with this shortened P2 still showed 10-fold decreased cleavage rate, reinforcing the fact that L5c forms a tertiary contact with an element other than P2.

Next, ideal solvent accessible (SAS) data and SAXS data were generated for 3D models and compared to the experimental data. The ideal SAS data of each model agrees well with each other except at the P2 region, while the footprinting data sits in between them at this region (Figure 4.13). In most of other regions of the *Bangia* ribozyme, ideal data from both models is similar to the real data. Differences are found at regions that are also observed to be different in the ideal and experimental data of the *Tetrahymena* ribozyme (not shown here). Therefore, those dissimilarities might come from the uncertainty of SASA calculation or limitations in the template *Tetrahymena* RNA. Analogously, the ideal SAXS data showed a smaller radius of gyration (R_g) than experimentally determined, which was also seen in the *Tetrahymena* ribozyme, indicating systematic variation inherent to the template model (data not shown). It is probably due to the more dynamic conformation of RNA molecules and thicker hydration shell around them in the SAXS experiment.

4.4 DISCUSSION

RNA structure is essential to its biological function. It has been described as hierarchical and modular. Secondary structures are highly stable and form in the absence of tertiary structures. Helices are then oriented specifically and connected by modular tertiary motifs in the three-dimensional structure. Nevertheless, our knowledge about this intricate system is still limited. Group I introns are good models for studying the architecture of functional RNAs since the 'peripheral' helices diverge largely in different subgroups, while they all have to wrap and buttress a similar catalytic core. We have probed the tertiary structure of a novel group IC1 intron from the *Bangia* rRNA via a catalytic activity assay, small angle X-ray scattering, hydroxyl radical footprinting, and computational modeling. We found that the *Bangia* group I ribozyme has most of the typical tertiary motifs that belong to this subgroup, which can be represented by the well-established *Tetrahymena* group IC1 ribozyme. Intriguingly, two novel structural elements in this ribozyme have been accommodated in the scaffold and provide examples for motif modularity in RNA structures.

The *Bangia* ribozyme has an extra stem-loop P5b.1, located the middle of the P5b stem. The entire P5abc element that extends from the P4-P6 domain is actually only present in certain intron subgroups, such as IC1 and IE (*72, 73*). In the *Tetrahymena* ribozyme, P5abc forms several tertiary contacts with the core and other peripheral elements including P5a-J4/5, P5b-J6a/6b and L5c-L2, which are important for the overall stabilization of the RNA. In the *Bangia* ribozyme, we revealed by mutagenesis and activity assay, that the L5b.1 instead of L5b is involved in a tertiary contact. This is possible since these two stem-loops are very close to each other, and both of them are

GNRA tetraloops that may form interactions between the J6a/6b receptor. Computational modeling also supports this hypothesis by providing a feasible local arrangement.

Another unique feature of the *Bangia* ribozyme is the short P8 stem that is capped by a GNRA tetraloop. Mutating the tetraloop into UUCG largely decrease the ribozyme activity, indicating the importance of this motif for RNA stability. A potential receptor of the P2 stem has been identified by comparative analysis and tested by activity assay. The single mutant of P2 has the same activity as the L8 UUCG mutant. More importantly, the P2-L8 double mutant also gives the same rate constant in the cleavage assay, strongly suggesting the correlation between these two sites. However, this finding introduces an interesting question as to which orientation of P2 and P8 stem would allow this novel contact. One possibility is that the *Bangia* ribozyme applies the same strategy as the *Tetrahymena* ribozyme since they are both group IC1 introns and have very similar peripheral elements. The other possibility is that the *Bangia* ribozyme may arrange P2 in the same way as the *Azoarcus* ribozyme, which is also reasonable because they both have a tertiary contact between P2 and P8. The only concern is that the L2 tetraloop docks on the P8 receptor in the *Azoarcus* (*88*), while L8 tetraloop docks on P2 receptor in the *Bangia*.

Computational modeling gave us an opportunity to model the 3D structure and evaluate both hypothesizes. In the model that applied the first strategy, P2 coaxially stacks on P2.1 and points to the P5abc domain. Notably, this arrangement exposes the putative receptor in the helix minor groove, and positions it close to the P8 stem; whereas the short P8 stem (three base-pairs) is highly restricted by its adjacent P3 helix and therefore puts the L8 tetraloop right next to the P2 receptor. In contrast, the putative P2

receptor is sequestered inside a helix when P2 stacks on P1. Indeed, the second model showed that the P8 is positioned in the same place as the first model due to the helical stacking with P3. This structural change puts the end of P2 near the L8 tetraloop and suggests a tertiary interaction between them. This alternative contact was less plausible because a mutant with shortened P2 acquainted the same activity as the wild-type ribozyme. Overall, the first model is considered as a more plausible 3D structure of the *Bangia* ribozyme.

All group I introns share a preserved active core, which is stabilized by a variety of peripheral helices in different subgroups. Cech has brought out a question whether the presence of a certain peripheral element is indicative of the existence of its corresponding partner, or it may be maintained during the evolution and can be used under different circumstance or controlled by protein factors in regulation (*16*). The structure of the *Bangia* group I ribozyme might address this inquiry. Analogous to the *Tetrahymena* ribozyme, it also puts the P2 stem in a stacked position with P2.1, which together wraps around the active site. Intriguingly, the presence of the short P8 stem and the L8 GNRA tetraloop enables a new contact that can help in positioning P2, while the other contact between L2-L5c that is responsible for stabilizing P2 in the *Tetrahymena* ribozyme has been lost. The helical elements P8 and P2 are prevalent in diverse subgroups, but they only contact each other in some cases (for instance, IA and IC3 group I introns). It is very unlikely that P8 is present in the *Bangia* ribozyme because of P2 or vice versa. A rational interpretation might be deduced from the 'soil' theory, that the P8 and P2 kept in group I introns evolved a contact between each other, which makes the other tertiary contact (L5c-L2) redundant and lost eventually in the *Bangia* group I intron. However,

investigation of more group I introns is necessary to establish an integral picture for RNA structure and evolution.

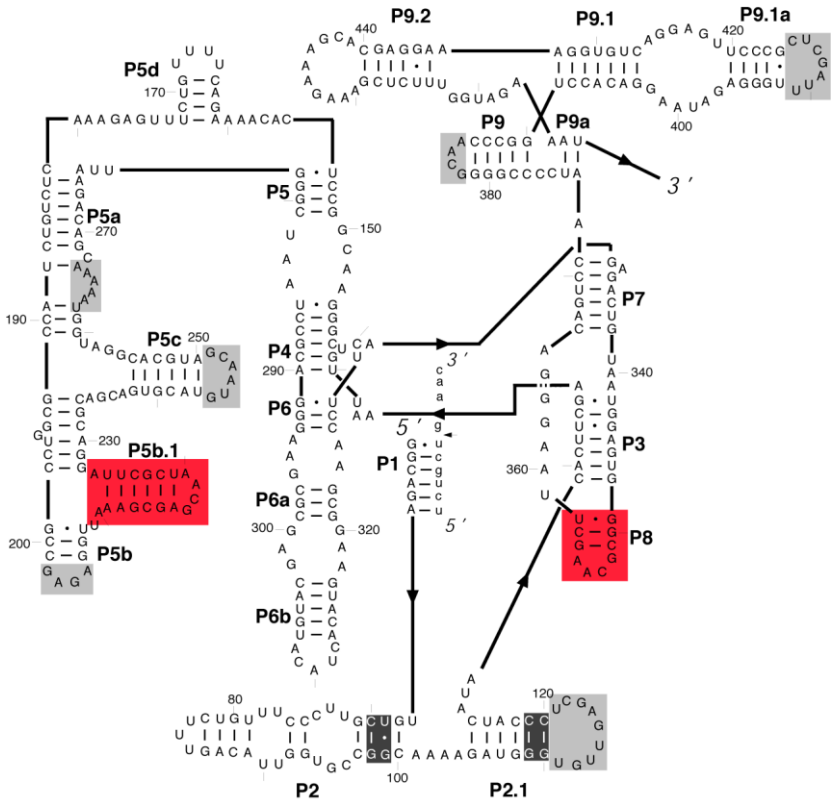

Figure 4.1: Secondary structure of the *Bangia* group I intron

Typical group IC1 tertiary elements are shaded in gray, and the two novel GCAA tetraloops (L8GCAA and L5b.1GCAA) are highlighted in red. Putative receptors in P2 and P2.1 are shaded in black.

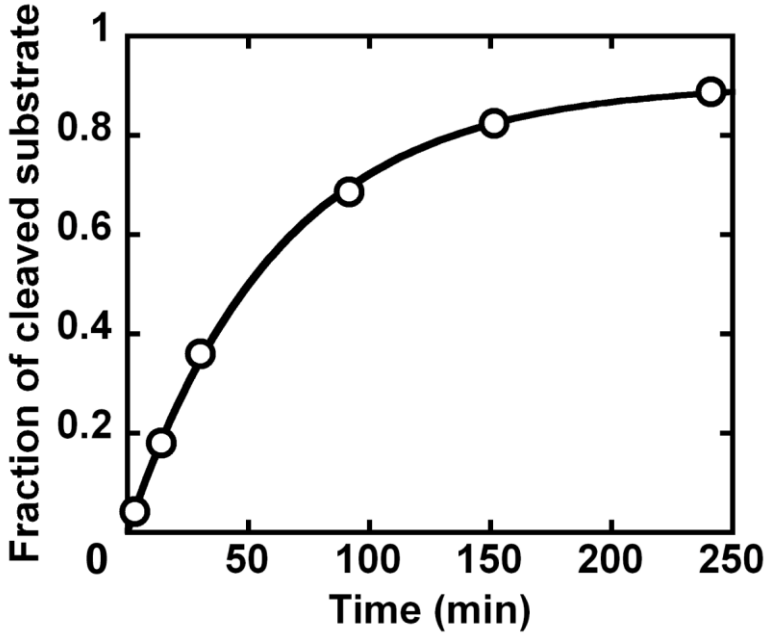

Figure 4.2: Measurement of catalytic activity of the wild-type *Bangia* ribozyme

RNA was preincubated at 50 ℃ for 30 minutes, then labeled substrate was added (25 ℃, pH7, 10mM Mg^{2+}, 50mM Na-MOPS, 500uM G, 400 nM ribozyme). The cleavage rate of the wild type ribozyme is 0.021 ± 0.005 min^{-1}.

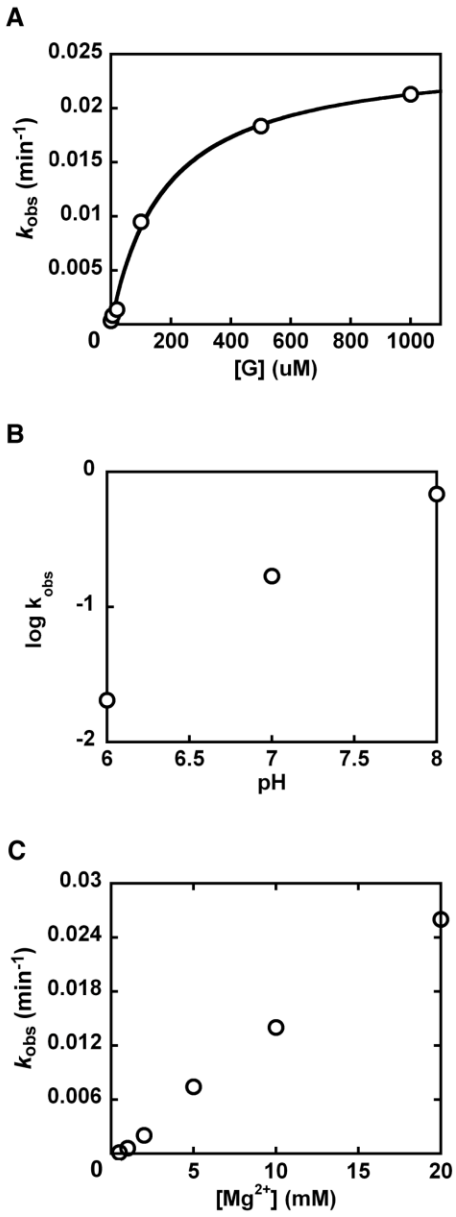

Figure 4.3: Varying conditions to confirm that the cleavage rate is rate-limiting in the catalytic activity assay

(A) Guanosine dependence of the *Bangia* ribozyme activity. The K_D is 180 nM. (B) pH dependence at 100mM Mg^{2+}. The cleavage rate increases log-linearly with pH. (C) Mg^{2+} dependence of the ribozyme activity.

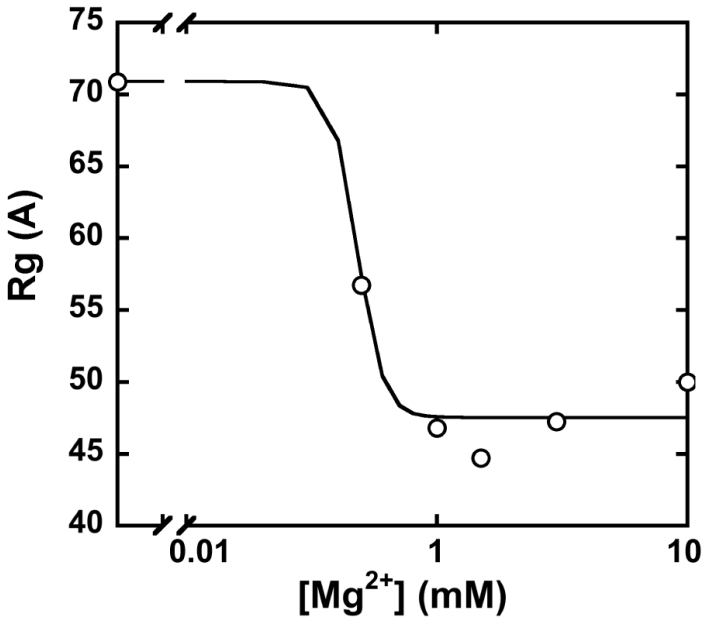

Figure 4.4 Mg^{2+} dependent folding of the *Bangia* ribozyme monitored by SAXS

The RNA (2 μM) was preincubated with varying Mg^{2+} at 50°C for 30 minutes, and then transferred to the measuring cuvette at 25°C. R_g drops from 71 Å to around 44 Å upon addition of Mg^{2+} ions. The $K_{1/2}$ is 0.5 mM Mg^{2+}, whereas the Hill coefficient is 8. Because the free Mg^{2+} concentration is lower than indicated in the plot (RNA molecules take up an estimated about 1 Mg^{2+}/nucleotide) (*31, 85*), the real $K_{1/2}$ is smaller than 0.48 mM. The increase of R_g from 1.5-10 mM Mg^{2+} is probably due to a small amount of aggregation.

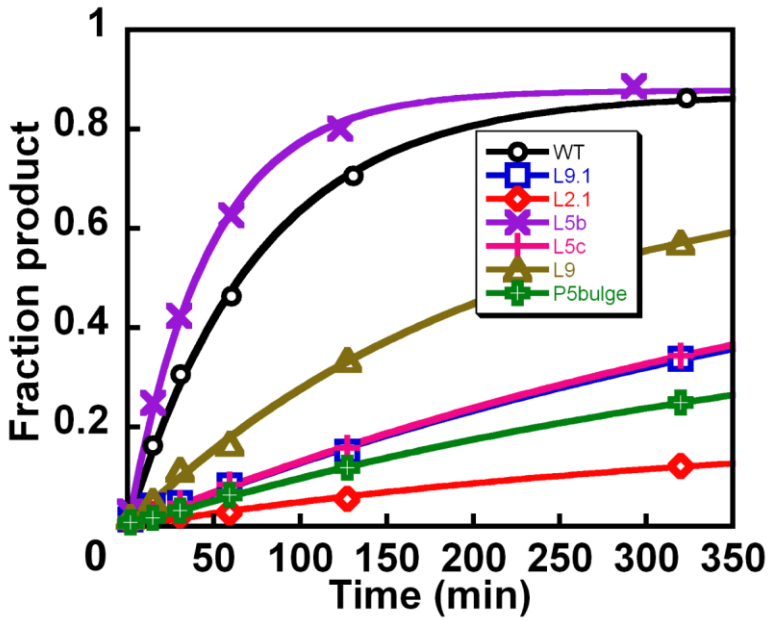

Figure 4.5: Effects of tertiary element mutations on catalytic activity

The cleavage activity of all mutant ribozymes were performed under standard conditions, and the results are plotted in different color as indicated in the legend. The data of rate constants are indicated in Table 4.1.

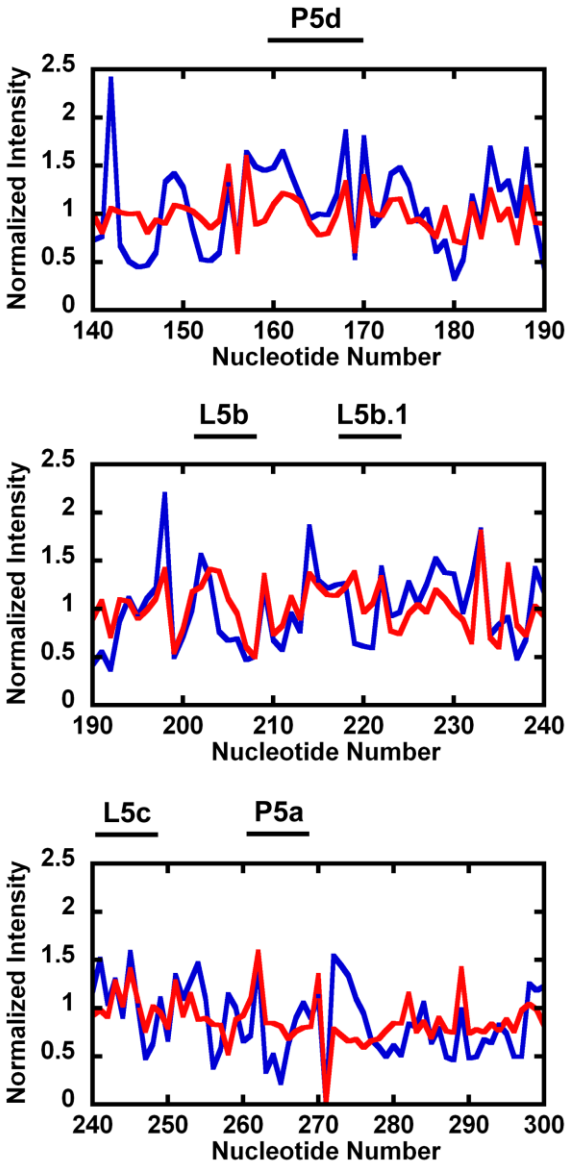

Figure 4.6: Comparison of folded *vs.* unfolded *Bangia* ribozyme by hydroxyl radical footprinting

The footprinting experiments were done to explore the solvent exposure of folded (10 mM Mg^{2+}) (blue) and unfolded (no Mg^{2+}) ribozymes (red). Only residues between 140 and 300 are shown here. Protection or exposure pattern are different between folded and unfolded ribozymes at all the tertiary elements (indicated by black lines on the top), even at the L5b loop which is predicted to not affect activity when knocked out.

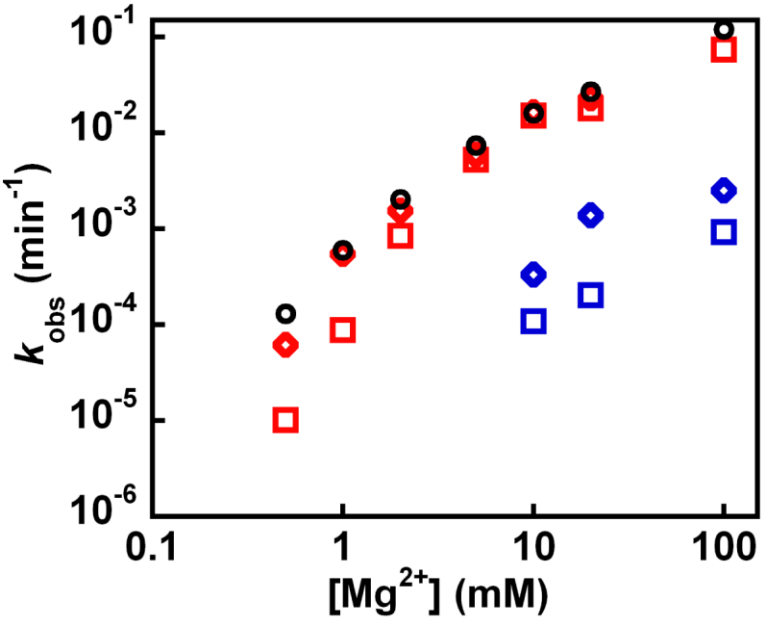

Figure 4.7: Mutations in the L8 and L5b.1 loops affect activity differently

Mutating the L8 GCAA tetraloop (blue) into UUCG (squares) and GCGA (diamonds) both largely decrease the activity at standard conditions (10 mM Mg^{2+}) and higher Mg^{2+}. On the other hand, mutating the L5b.1 GCAA tetraloop (red) into UUCG (squares) only showed effect at and below 1 mM Mg^{2+}, while replacing it by GCGA (diamonds) barely show any effect even below 1 mM Mg^{2+}.

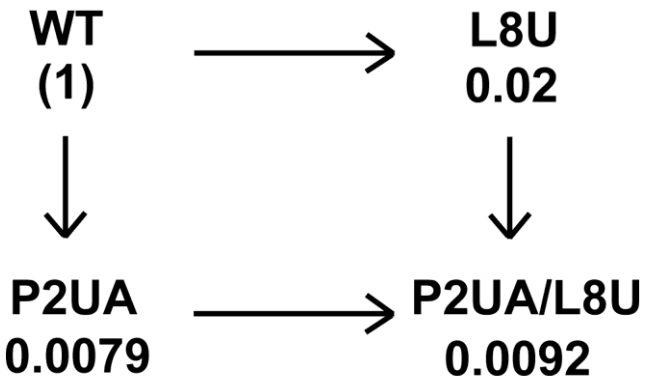

Figure 4.8: Double mutant cycle of L8UUCG mutant and P2UA mutant

Activity of each mutant was measured under standard conditions (25 ℃, pH7, 10mM Mg^{2+}, 50mM Na-MOPS, 500uM guanosine, 400 nM ribozyme). The relative rate constant for each mutant was normalized against that of wild-type ribozyme measured side-by-side. The data are also shown in Table 4.2.

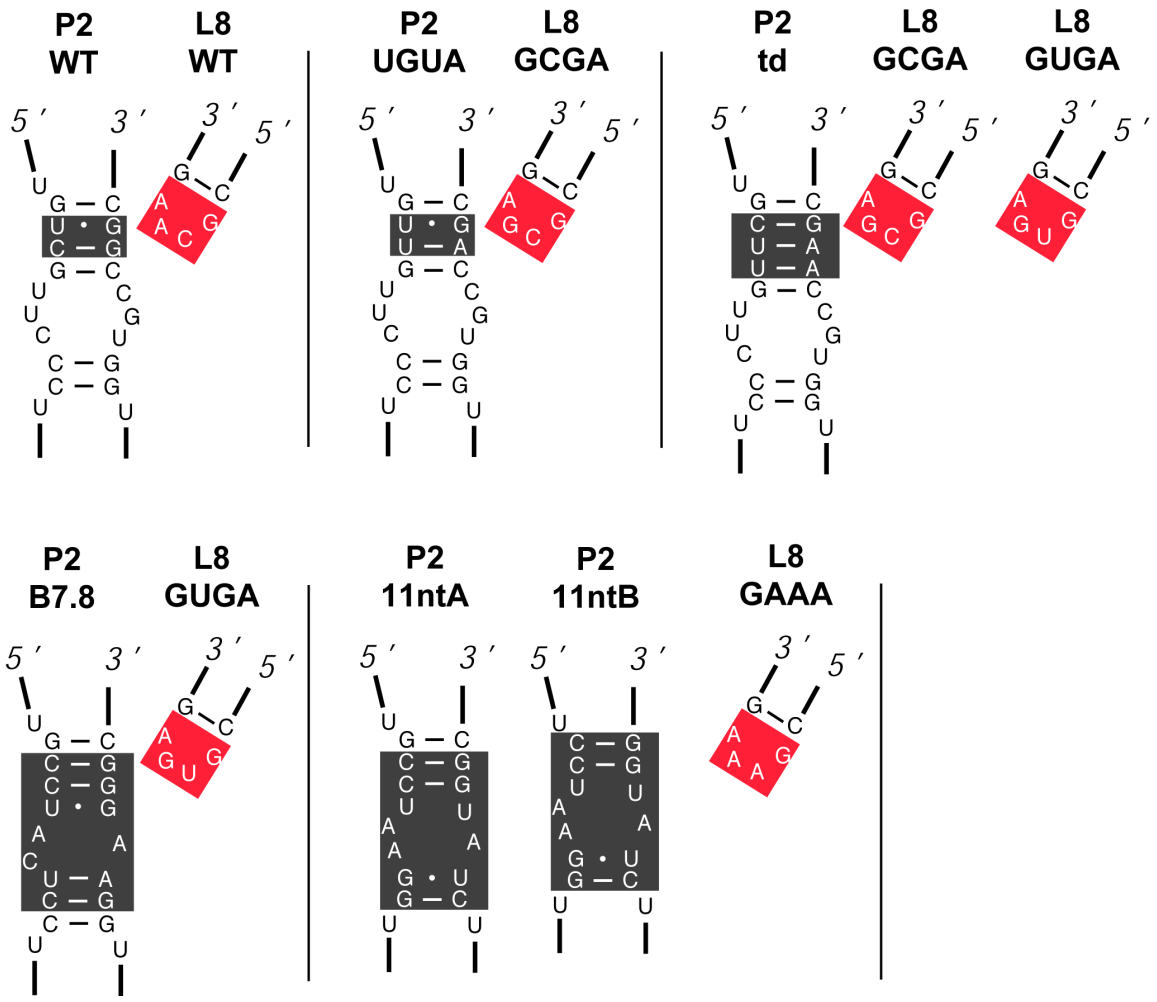

Figure 4.9: Design of rescue mutants to confirm the contact between the L8 tetraloop and the P2 receptor

The mutated sequences in the L8 loop are shaded in red while the corresponding design in P2 is shaded in black. The design of receptors for different types of GNRA tetraloops are based on previous *in vitro* selection studies (*163*).

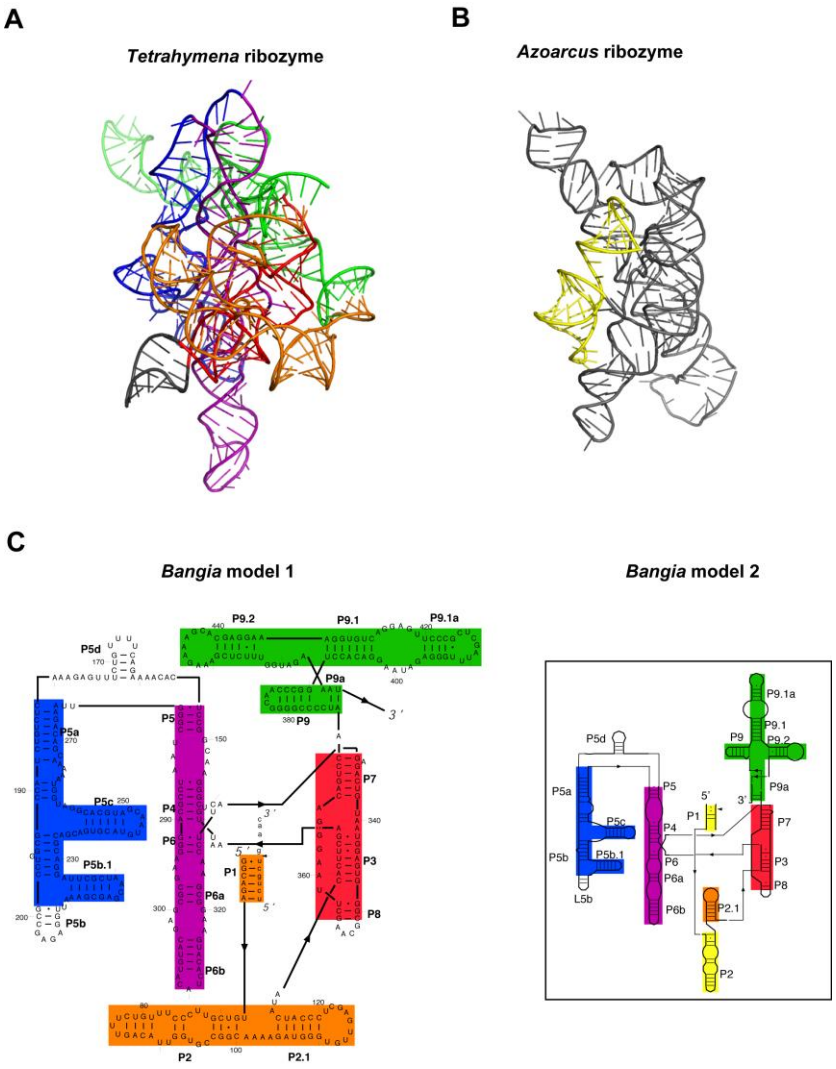

Figure 4.10: Threading of the *Bangia* ribozyme on the structures of the *Tetrahymena* and *Azoarcus* group I ribozymes

(A) 3D structure model of the *Tetrahymena* ribozyme (*76*). Helices not used in modeling are colored in grey. (B) Crystal structure of the *Azoarcus* ribozyme (*87*, *88*). P1-P2, which is used as a template for model 2, is colored in yellow (C) Secondary structure of the *Bangia* ribozyme. Each secondary element of the *Bangia* RNA in model 1 and model 2 is colored according to the color of the template in plot A and B.

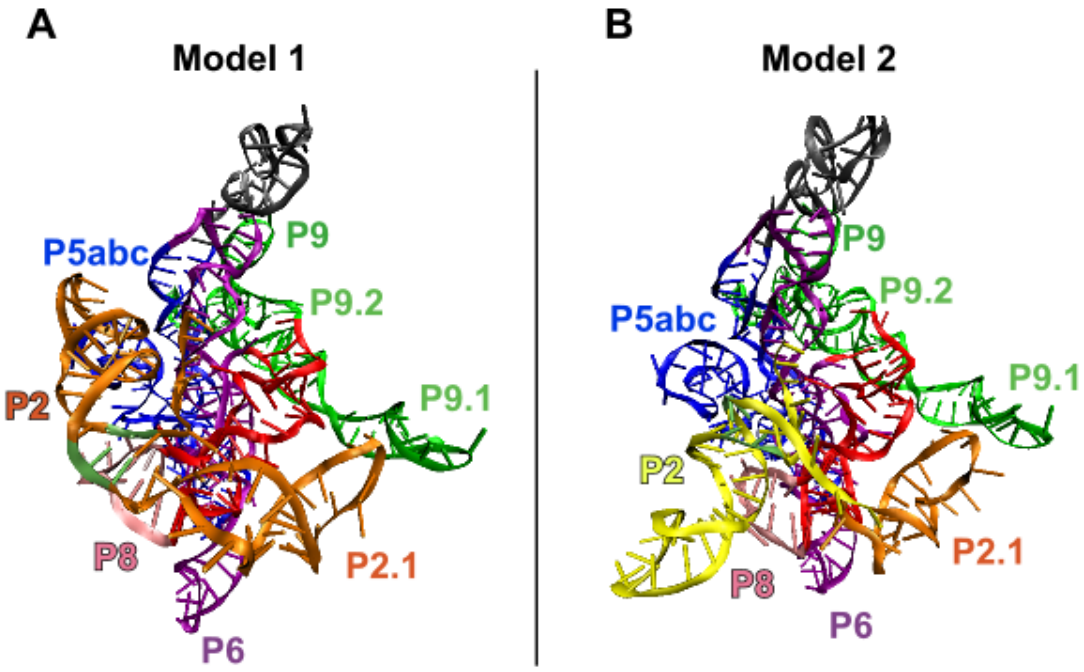

Figure 4.11: Computational models of *Bangia* ribozyme generated by RNABuilder

(A) The model was generated by using a template with mostly *Tetrahymena* ribozyme elements and P1-P2 from the *Azoarcus* ribozyme. Helices nomenclature is indicated. The color of each duplex is correspondence to the template shown in Figure 4.10. Additionally, in this figure and panel B, the putative receptor site in P2 helix is highlighted in lime, and the P8 helix is highlighted in pink. (B) Model 2 generated by using a template with most of the *Tetrahymena* ribozyme elements but P1-P2 from the *Azoarcus* ribozyme.

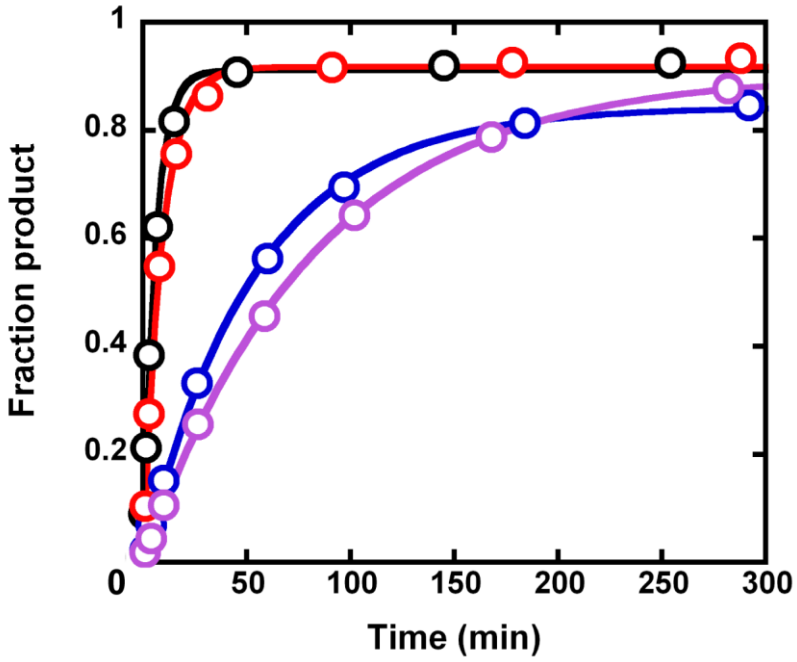

Figure 4.12: Cleavage activity of P2 shortening mutants

All wild-type and mutant ribozymes were preincubated at 50 ℃, 10 mM Mg^{2+}, 50 mM Na-MPOS, pH 7 for 1 hour, and mixed with trace radio-labeled substrate, 450 μM guanosine and 100 mM Mg^{2+} to start the cleavage assay. The rate constant of P2 mutant (with only 4 bp in the P2 stem) is similar to the wild-type ribozyme: 0.11 min^{-1} (red) and 0.17 min^{-1} (black), respectively. The double mutant P2/L5c gives a similar rate (0.012 min^{-1}, purple) as the L5c mutant (0.019 min^{-1}, blue).

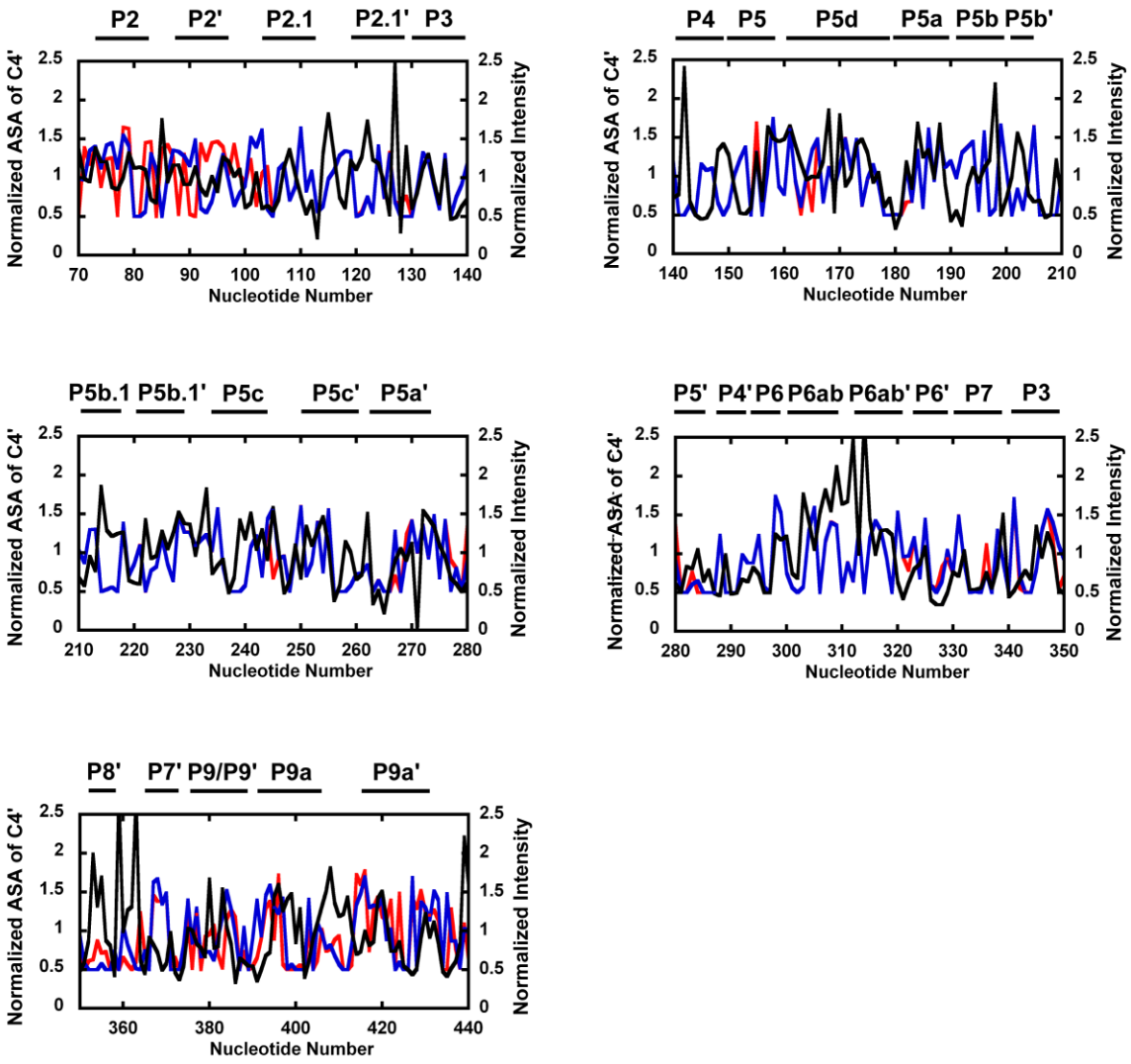

Figure 4.13: Comparison between the ideal solvent accessibility data of 3D models and the hydroxyl footprinting data

Hydroxyl radical footprinting data (black) of each residue is normalized by the average intensity. Simulated solvent access area (SAS) of C4' atoms in model 1 (red) and model 2 (blue) are normalized by the same extent to be comparable to the normalized footprinting data. The solvent protection pattern but not values of each residue are compared between the simulated data and the experimental data.

	K_d^E	K_d^{IGS} (calculated)
25 ℃	3 nM	4 nM
37 ℃	21 nM	210 nM

Table 4.1: Tertiary stabilization from the *Bangia* ribozyme on the substrate

The substrate sequence is UCUGGUGAAC from 5' to 3'. The K_d^E was measured under saturating cofactor guanosine (450 μM G, 10 mM $MgCl_2$, 50 mM Na-MOPS, pH 7.0) and varying ribozyme concentrations (20 – 200 nM). Value for k_{cat}/K_m was also obtained in this assay, giving 3.6 x 10^6 M^{-1} min^{-1} at 25 ℃, and 1.9 x 10^6 M^{-1} min^{-1} at 37 ℃. Calculated K_d^{IGS} was obtained by the nearest neighbor rule between substrate and IGS (GGCAGA from 5' to 3'), according to previous work (*158*).

Mutants	WT seq.	Variant seq.	k_{obs}, $(min^{-1})^1$	k_{obs}, rel[2]
WT	\	\	1.3×10^{-2}	(1)
L9.1	UUAGCUC	UUCG	1.4×10^{-3}	0.11
L2.1	UCUUGAGC	UUCG	4.5×10^{-4}	0.03
L5b	GAGA	UUCG	2.1×10^{-2}	1.3
L5c	GUAACG	UUCG	1.4×10^{-3}	0.11
L9	GCAA	UUCG	3.2×10^{-3}	0.21
P5a	AAAAA	UUUUU	9.9×10^{-4}	0.07

Table 4.2: Activity measurements for tertiary contact disruption mutants.

[1]25 ℃, pH7, 10 mM Mg^{2+}, 50mM Na-MOPS, 500 μM guanosine

[2]The reaction rate constant for each variant is divided by the rate constant determined for the wild type ribozyme in the same experiment ($0.013min^{-1}$). The activities of most of the mutants have also been measured at 0.5 mM Mg^{2+} and gave similar relative $k_{obs,}$ rel (data not shown).

Mutants	**k_{obs}, (min^{-1})[1] (10 mM Mg^{2+})**	**k_{obs}, (min^{-1}) (100 mM Mg^{2+})**	**k_{obs}, rel[2] (10 mM Mg^{2+})**
WT	1.4×10^{-2}	0.25	(1)
P2UA	1.1×10^{-4}	1.1×10^{-3}	0.0079
L8	2.9×10^{-4}	0.8×10^{-3}	0.02
P2UA/L8	1.3×10^{-4}	1.1×10^{-3}	0.0092
L5c	1.4×10^{-3}	2.2×10^{-2}	0.1
P2UA/L5c	2.0×10^{-5}	1.7×10^{-4}	0.0014

Table 4.3: Activity measurements for L8/P2 double mutant and L5/P2 double mutant

The P2UA mutant changes the two consecutive GC base pairs in the P2 region into U-A base pairs. The L8 and L5c mutants change the tetraloops at L8 and L5c into UUCG tetraloop, the double mutants P2/L8 and P2/L5c were mutated on both sides.

[1]25 ℃, pH7, 10mM Mg^{2+}, 50mM Na-MOPS, 500 μM guanosine

[2]The reaction rate constant for each variant is divided by the rate constant determined for the wild type ribozyme at 10 mM Mg^{2+}.

Chapter 5: The misfolding of a *Bangia* group I ribozyme follows a rapid collapse

5.1 INTRODUCTION

Structured RNAs, as proteins, also need to specify a native conformation to be functional. From tRNA to ribosomal RNA, all of them are facing folding problems (*34, 35*). But the strategy employed by RNAs is distinct from that of proteins owing to their distinct physical properties. The polymer of protein is constituted of at least 20 different monomers with extruding their side chains outwards for interaction, whereas RNA only has 4 different side chains that are sequestered in the helix. Therefore it is much more difficult for RNA to specify a single native form (*7, 9*). Indeed, numerous *in vitro* assays have shown that misfolding prevails in structured RNAs, though the molecular nature of intermediates varies in different models.

A good model to study the misfolding of RNA molecules is the self-splicing group I intron (*70*). It was engineered into a ribozyme version that mimics only the first step of splicing to make the interpretation between structure and function simpler and more straightforward (*74*). Among those group I RNAs, *Tetrahymena* group IC1 ribozyme has been investigated extensively for its folding pathway and intermediates. Previous work showed that the *Tetrahymena* ribozyme folds into a compact intermediate in milliseconds (*31, 79*), some of which then partitions into the native structure while the majority forms a misfolded conformation with a rate around 1 min^{-1} (*53, 58, 59, 80, 81, 83*). The misfolded species is structurally similar to the native conformation, but refolds into the

native form very slowly and requires extensive unfolding. A model has been proposed based on these results, proposing that there is topological difference between the misfolded and native species in the *Tetrahymena* ribozyme (*86*). This topological difference arises due to a nonnative duplex, alt P3 in the core during an early stage of folding. This mispaired region bias the nonnative topology inside the core of ribozyme which is stabilized by the native long-range tertiary contacts, but AltP3 seems to transform to the native P3 duplex in the final misfolded structure. Misfolded intermediates of other group I ribozymes have also been studied to extend our understanding in the structural features that lead to misfolding. For example, a group IE ribozyme from *Candida* that possesses almost the same peripheral elements as *Tetrahymena* ribozyme also populates misfolded species which may originate from a nonnative P3P7 structure as well (*97, 98*). In contrast, a small group IC3 ribozyme from *Azoarcus* that lacks most of the peripheral elements folds into a catalytically active structure without accumulating misfolded intermediates under specific conditions (*100, 101*), reassuring that peripheral elements may stabilize misfolded structure.

Here we explore the misfolding of a novel group IC1 ribozyme from Australia *Bangia* (red algae), the tertiary structure of which has been described in the last chapter. It has two novel tetraloops that are only present in *Bangia* introns, while its core sequence and peripheral elements are conserved in the IC1 subgroup. Here we present the results of small angle X-ray scattering that showed the existence of the nonnative conformation of the *Bangia* ribozyme upon adding metal ions. Additional hydroxyl radical footprinting and catalytic activity assay revealed the rates of compaction and formation of native structure, respectively. These preliminary investigations not only unveil the folding

pathway of the *Bangia* ribozyme but also exemplify a general rule of group I intron folding.

5.2 METHODS

5.2.1 Time-resolved small angle X-ray scattering

The static SAXS measurements have been described in Section 4.2.4 and reference (*166*). The ribozyme (2 μM) was preincubated under varying conditions (25–50 °C, 1-50 mM Mg^{2+}, 0-20 h). In the time-resolved SAXS experiments, RNA was mixed with Mg^{2+} immediately before the beam time using a remote controlled mixer, and folding was measured on the second time-scale (*31*).

5.2.2. Folding of the *Bangia* ribozyme monitored by hydroxyl radical footprinting

The hydroxyl radical footprinting was carried analogously to the process described in Section 2.2.6. The wild type *Bangia* ribozyme was 5' radio-labeled and folded at different temperatures (pH 7.0, 50mM Na-MOPS, 10mM Mg^{2+}, 25-50°C). The control with unfolded RNA was also generated in parallel. Fe(II)-EDTA footprinting reagents (2 mM $(NH_4)_2Fe(II)(SO_4)_2$, 2.5 mM Na-EDTA, 6mM Na-ascorbate) were mixed with each sample for 10 min at 25°C. Footprinting reactions were quenched in Fe(II)-EDTA quench solution. Only the folding pattern of the 5' half (1-200 nt) of the ribozyme was monitored by footprinting, whereas the 3' half will be conducted in the future.

5.2.3 Activity assay to follow the formation of native ribozyme

At varying temperatures (25 °C, 37°C or 50 °C), wild-type *Bangia* ribozyme (200 nM) was incubated at pH 7.0, 50mM Na-MOPS and 10mM Mg^{2+}. Aliquots were quenched at various times and incubated with radio-labeled substrate and a 2-fold excess of cold substrate for 3 minutes. Guanosine, 100 mM Mg^{2+} and 20-fold excess cold substrate were then added to start the cleavage reaction, which was stopped by EDTA solution after various times. The products were run on 20% acrylamide/7M urea denaturing gel and quantitated by Phosphorimager.

5.3 RESULTS

5.3.1 Small angle X-ray scattering revealed a trapped misfolded construct

In the previous work, we have shown that the *Bangia* ribozyme folds into a globular structure upon addition of Mg^{2+} ions, with a radius of gyration of 44 Å measured at 1.5 mM Mg^{2+}. However, this does not reflect the folding intermediates that may accumulate or transiently exist. Time resolved SAXS experiment has previously showed that *Tetrahymena* ribozyme rapidly collapsed to an intermediate from the unfolded state, as indicated by a drop in the radius of gyration (*31, 85*). For the *Bangia* ribozyme, we mixed the RNA (2 μM) with metal ions (10 mM Mg^{2+}) and monitored its folding using stopped flow equipment. The result showed that within 10 seconds the R_g value decreases from 71 Å to about 52-55 Å, which is close to the value obtained in previous SAXS measurements at 10 mM Mg^{2+} (Figure 5.1); indicating the *Bangia* ribozyme also compacts into a folded state very rapidly. However, it is not clear so far whether this is an

intermediate or the native form. We then varied the conditions since the intermediates may have different properties than the native ribozyme under certain circumstances. We incubated the ribozyme at different Mg^{2+} concentrations and temperature, and followed the folding via static SAXS. At 50 °C and 10 mM Mg^{2+}, the R_g value decreases to around 52 Å within a minute; but when the Mg^{2+} is increased to 50 mM Mg^{2+}, the R_g value decreases from 61 Å to 56 Å within half an hour, suggesting the existence of intermediates (Figure 5.2). We have also decreased the temperature to 37 °C and 25°C at 10 mM Mg^{2+}, and a trapped intermediate with a R_g of 56 Å persists for at least 2 hours and 20 hours, respectively (data not shown). Therefore, all of the above results indicated that while the native conformation of the *Bangia* ribozyme is around 51 Å at room temperature and 10 mM Mg^{2+}, there is at least one misfolded species with a size of approximately 56 Å. This misfolded conformer was very stable and refolded to the native form slowly under this condition. This number is a little different than determined in time-resolved measurements, probably due to the differences in experiment set-up and errors.

5.3.2 The folding of the *Bangia* ribozyme involves multiple intermediates

To further investigate the properties of the intermediates and the folding pathway of the *Bangia* ribozyme, hydroxyl radical footprinting assay was performed to follow the transitions from intermediates to the native conformation. RNA was incubated at 10 mM Mg^{2+} and varying temperature; aliquots were taken out at various times to mix with hydroxyl radicals for cleavage. Since the compaction of the unfolded state was

accomplished within a minute, the process followed by footprinting assay here was only the transition from the compact intermediates to the native ribozyme.

The protection patterns of different regions of the *Bangia* ribozyme were then analyzed individually, and most of them showed a two-phase transition (25 °C, 10 mM Mg^{2+}). The variation between the rate constants of each phase was within 2- to 3- fold in different regions (e.g. P2, P2.1, P5abc) (Figure 5.3). Therefore, it is very likely then there are at least two intermediates before the native conformation in the folding pathway, and even though they are all compact, the local structures are different from each other. The first intermediate (I_1) folds to the second intermediate (I_2) with a rate constant of 0.05-0.09 min^{-1}, followed by the folding from I_2 to the native ribozyme at 0.002-0.005 min^{-1} (Table 5.1). When we raised the temperature to 37°C, folding was accelerated in both phases by at least 10-fold, indicating that folding of these misfolded structures is dependent on temperature. However, only one phase, with a rate of 0.5-1 min^{-1}, has been observed under 50 °C, possibly because the fast phase is too short to be followed by manual pipetting. Further dissection of these intermediates may require the acquisition of Mg^{2+} concentration dependent folding properties.

5.3.3 Formation of native ribozyme monitored by catalytic activity assay

Although the SAXS and footprinting experiments have shown compaction of the *Bangia* ribozyme in multiple steps, the formation of native ribozyme can only be reflected by catalytic activity assay. Since the cleavage activity of the *Bangia* ribozyme is quite slow (0.016 min^{-1} at 10 mM Mg^{2+} and 25 °C), we varied the conditions to make the fraction of native ribozyme conspicuous over time. The ribozyme was first incubated at

25 °C, 10 mM Mg^{2+}, 50 mM Na-MOPS, pH 7; aliquots were then taken out at intervals to mix with radio-labeled substrate and a 2-fold excess of cold substrate (Figure 5.4A). The 2-fold excess substrate was added in case substrate binds to the native ribozyme and intermediates at different rates, since single turnover assay can only be used on the assumption that those rates are the same. After allowing 3 minutes for complete binding between the ribozyme and substrate; cofactor guanosine (G), additional Mg^{2+} (100 mM final) and 20-fold excess cold substrate were added to start the cleavage reaction. Mg^{2+} was added to increase the cleavage rate of the *Bangia* ribozyme; while 20-fold excess cold substrate was added to make sure the ribozyme did not bind to the radio-labeled substrate again after the product release. Both strategies were applied to ensure a burst can be shown to reflect the fraction of native ribozyme during the folding process.

Indeed, a burst followed by a slow phase was observed. The burst amplitude reflects the fraction of native ribozyme that cleaves substrate, and the slow phase rate represents product release. We then plotted burst amplitude against the time of incubation to attain the folding rate to the native conformation (Figure 5.4B). The rate constant (0.0057 min^{-1}) was similar to the observed second phase rate constant in multiple regions of the ribozyme via footprinting (Table 5.1). The results indicated that none of those two intermediates have detectable activity under the standard conditions and only the slow step from I_2 to the native ribozyme was monitored by activity assay. Interestingly, there is also a small fraction of the *Bangia* ribozyme forms to the native conformation within a minute, as shown in the plot (Figure 5.4B), suggesting a pathway similar to the *Tetrahymena* ribozyme that partitions to the functional species in an early step of the pathway before the majority becomes trapped in intermediate structures.

The transition from the intermediate I_2 to the native form has been further tested analogously to refolding studies of the misfolded species in the *Tetrahymena* ribozyme, although the rate in the *Bangia* ribozyme is 20 fold faster. We first investigated whether the fast folding corresponded to less extensive unfolding during this process. Increased concentrations of urea were added during the measurement of folding from I_2 to native ribozyme. The *m* value obtained is -0.84 kcal mol^{-1} M^{-1} (Figure 5.5), which is almost half of that of the *Tetrahymena* ribozyme (*86*). It suggested that perhaps not all the long-range tertiary contacts need to be disrupted during the transition from I_2 to N. Indeed, when we mutated the L9.1 loop and the P5 bulge, no obvious acceleration was observed (data not shown). However, absence of extensive unfolding needs to be confirmed in future studies and more mutants need to be tested.

5.4 DISCUSSION AND FUTURE PLAN

Group I introns have been used as model systems to study structure and folding of functional RNAs for decades. Studies of the *Tetrahymena* group I ribozyme have provided extensive insights into RNA folding pathways and misfolded intermediates. We wanted to extend the investigation to more group I RNAs to establish the presence of a general rule in RNA folding may be applied in the future. The *Bangia* ribozyme which belongs to the same IC1 subgroup as the *Tetrahymena* ribozyme has almost the same secondary structure as the *Tetrahymena* ribozyme but two different motifs that may form different tertiary contacts. The folding pathway of the *Bangia* ribozyme therefore allows us to study novel folding features of group I RNAs, while building on the existence knowledge available for the *Tetrahymena* ribozyme.

We first monitored the compaction of the *Bangia* ribozyme by time-resolved SAXS, and found that this ribozyme also collapses into a compact form within less than a minute, similar to the *Tetrahymena* ribozyme. To answer the question whether this molecule represents the native structure, we varied the temperature and Mg^{2+} and found that a more extended intermediate is populated at higher Mg^{2+}. Hydroxyl radical footprinting assay further revealed the existence of more than one misfolded intermediate, since there are two kinetic steps in the folding of local region. Those two steps can either represent two inconvertible intermediates that convert to the native ribozyme with different rates or two consecutive intermediates that fold sequentially into the native form.

The results from the activity assay favored the second model because formation of the native state under the same conditions only has one kinetic step, which has the same rate constant as the slow phase in the footprinting assay. The active state of this group I ribozyme is formed from trapped intermediates with a rate of 0.0057 min^{-1}, about 20 fold as fast as refolding from the misfolded specie to the native form in the *Tetrahymena* ribozyme. This rate is coincidentally the same as that of the transition from I_2 in the solvent radical protection, illustrating that both footprinting and activity assay have followed the same transition, and the two intermediates found in the footprinting assay are consecutive in the folding pathway. In addition, the activity assay revealed a small fraction of native ribozyme in the early stage before formation of intermediates, suggesting a partitioning step similar to the *Tetrahymena* ribozyme. In summary, our models suggested that unfolded *Bangia* ribozyme rapidly collapses to a transiently populated intermediate, five percent of which then partitions to the native state, while the

remaining fraction goes to I_1 and rapidly folds into I_2. The refolding from I_2 is faster than that from M to the native state in the *Tetrahymena* ribozyme, and requires less unfolding. A plausible model is that the different tertiary arrangement in the *Bangia* ribozyme makes the transition between intermediates and native form easier.

Folded region	Residue number	Refolding rate (min^{-1})		
		25°C	37°C	50°C
P2	95-97	ND	0.68/0.040	ND
L2.1	113	0.079/0.0019	ND	ND
J3/4	137-139	ND	ND	1.0
J4/5	144-147	ND	ND	0.64
P5d	161-162	ND	0.81/0.064	ND
P5d	173-175	0.088/0.0016	0.45/0.068	1.1
P5b.1	219-220	0.090/0.0039	0.042	1.3
P5c	249	ND	ND	0.47

Table 5.1: Folding of individual regions in the *Bangia* ribozyme mmonitored by hydroxyl radical footprinting

At 25°C and 37°C, the folding of most of regions in the *Bangia* ribozyme showed a rapid phase following by a slow phase. The averaged rate constants for both phases are presented in the table and separated by a slash line. At higher temperature, especially 50 °C, only single-phase folding was observed and the averaged rate constants are shown. ND: not determined.

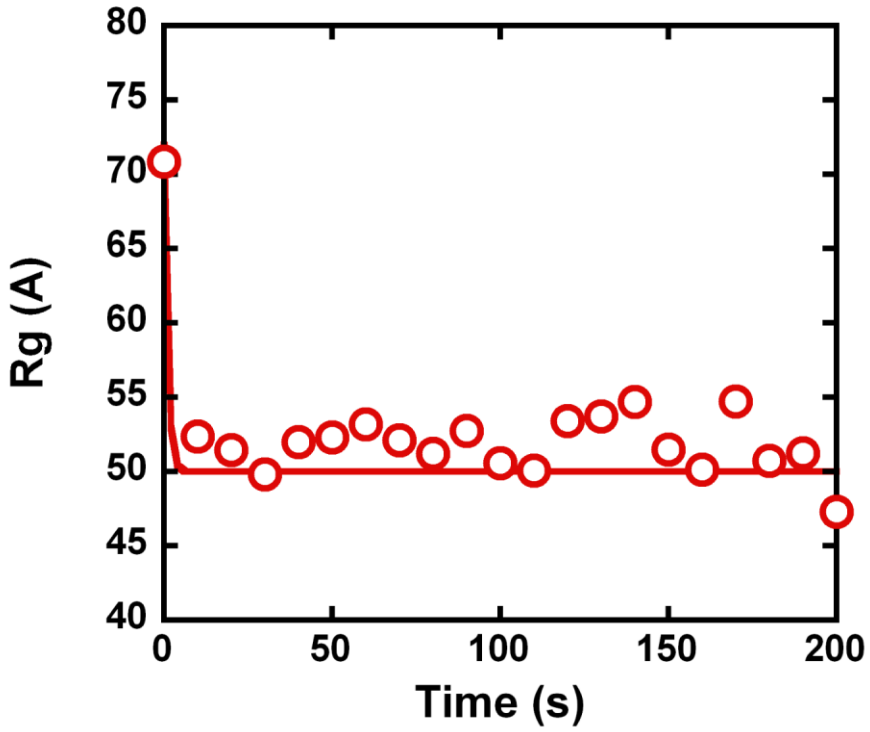

Figure 5.1: Compaction of the *Bangia* ribozyme monitored by SAXS

The ribozyme was mixed with 10 mM Mg^{2+} at 25 ℃. Radius of gyration (R_g) decreased to around 51 Å before the first time point (10 s), indicating a rate constant greater than 8 min^{-1}.

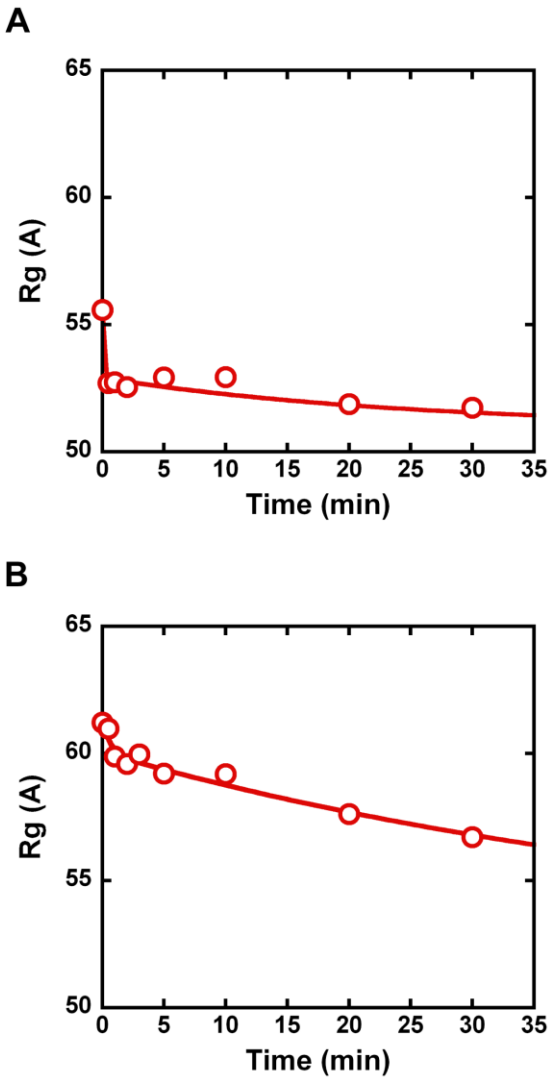

Figure 5.2: SAXS reveals partially folded intermediates of the *Bangia* ribozyme

(A) Time course of folding at 50 ℃ and 10 mM Mg^{2+}. Ribozyme (2 μM) was incubated at 50 ℃ for various time, and moved to the static cuvette at room temperature and measured by X-ray beam. The time point 0 is the R_g of ribozyme incubated at 10 mM Mg^{2+} and 25 ℃ briefly. After moving the RNA to 50 ℃, the RNA collapsed rapidly with a rate larger than 10 min^{-1}. (B) Time course of folding at 50 ℃ and 50 mM Mg^{2+}. The incubation at time 0 was performed as described in panel A, but the R_g value was much greater (61 Å), and it decreased to 56 Å after 30 minutes.

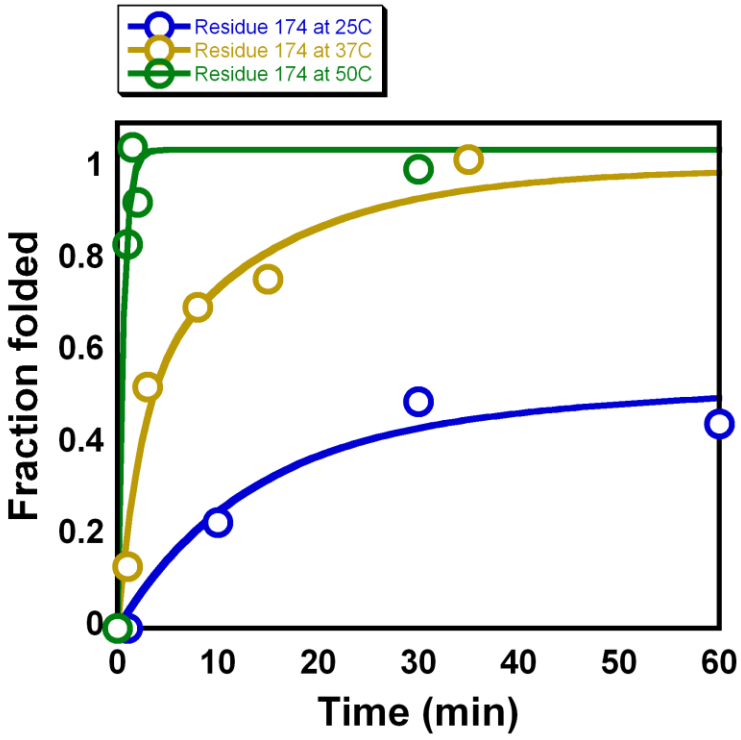

Figure 5.3: Time-dependent protection of residue 174 monitored by hydroxyl radical footprinting

Fraction of folding for individual residue 174 was determined by normalizing the difference between the maximum value (pre-folded at 50 ℃ for 30 min) and the minimum value (unfolded ensemble). At 25 ℃ and 37 ℃, the folding curves were double exponential curve (see the rate constants in Table 5.1), whereas it was single exponential curve at 50 ℃..

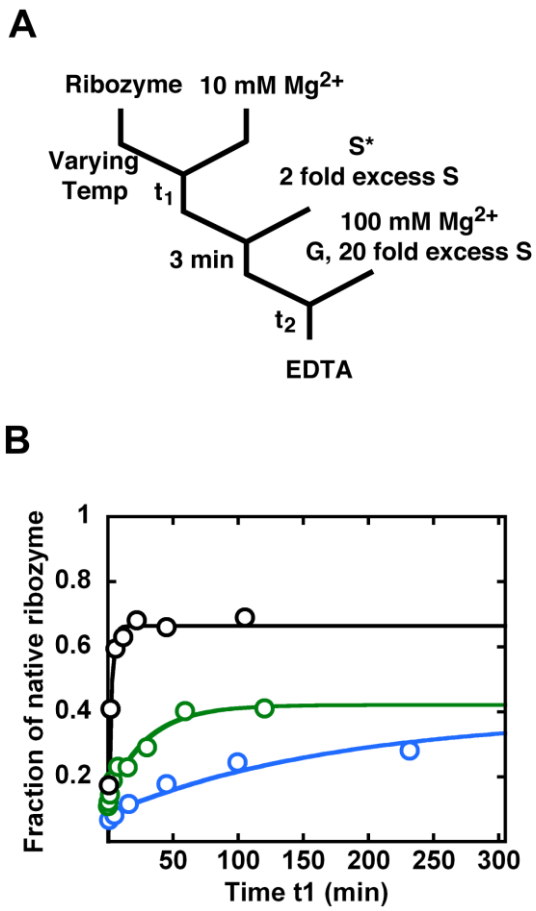

Figure 5.4: Formation of the native ribozyme monitored by activity assays

(A) The experimental design of activity assay to monitor the folding of the native *Bangia* ribozyme. (B) Refolding of the *Bangia* ribozyme at different temperatures. The rate is dependent on temperature: 0.0057 min^{-1} at 25 ℃ (blue), 0.034 min^{-1} at 37 ℃ (green), 0.33 min^{-1} at 50 ℃ (black).

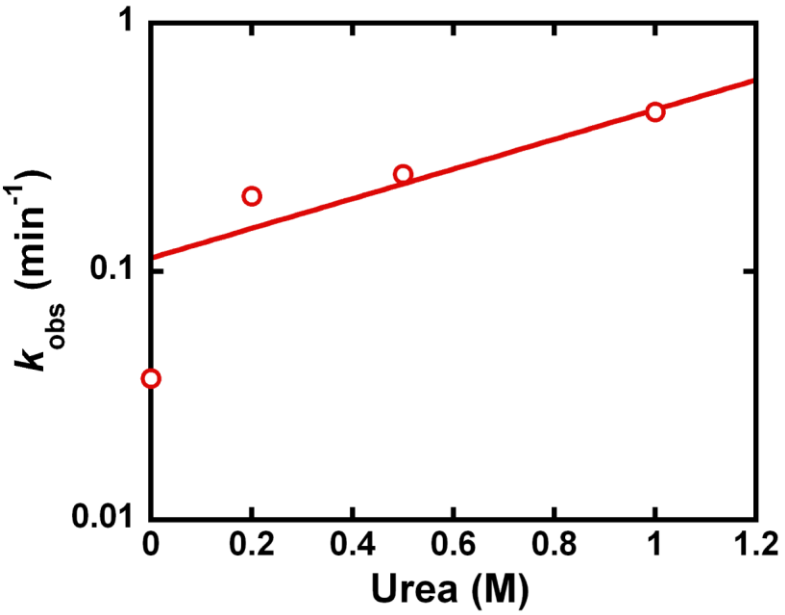

Figure 5.5: Urea dependence of refolding kinetics of the *Bangia* ribozyme

The measurements were performed under standard catalytic activity conditions (25 ℃, 100 mM Mg^{2+}, 50 mM Na-MOPS, and pH 7.0) and urea was added during folding and diluted in the cleavage assay. Plotting the folding rate against the urea concentration gave an *m* value of -0.84 kcal mol^{-1} M^{-1}.

Appendix

A.1 DMS FOOTPRINTING

DMS footprinting was performed essentially as previously described (*141*). Reactions contained 2 μM L-21/ScaI ribozyme in 25 μl of 50 mM Na–Mops, pH 7.0, and 10 mM Mg^{2+} solution at 15 °C. To give maximal population of I_{trap} for footprinting, we folded the ribozyme at 15 °C in the presence of Mg^{2+} for 1 min, followed by addition of DMS (1 μl of 16% DMS in ethanol) to a final concentration of 0.64%. DMS was allowed to react with the ribozyme for 1 min. The previously determined rate constant of 0.15 min^{-1} for decay of I_{trap} to M and N at 15 °C indicates that 85% of the population remains in the I_{trap} form at the time of DMS addition and 75% remains at the end of DMS treatment (see Figure 3.2) (*58*). DMS reactions were quenched and processed as previously described and analyzed using SAFA (*86, 139*). Band intensity values were normalized by dividing by the intensity of the fully extended product, and results from two to four independent determinations were averaged. The final intensity values were multiplied by an arbitrary scaling factor of 800 to allow comparison with hydroxyl radical results on the same scale, as shown in Figure 3.3 to Figure 3.5. In these figures, asterisks denote the most prominent changes in DMS footprinting, as defined by a change in the intensity values of at least 0.2 and a relative change of at least 30%. This analysis avoids nucleotides that are modified by DMS so strongly that they change in reactivity from one ribozyme form to another by a large amount and are therefore colored strongly, but whose changes are small relative to their total signal and therefore are likely to be less significant.

A.2 TRANSITION FROM ITRAP TO N REQUIRES LESS STRUCTURAL DISRUPTION THAN FROM M

Previous research demonstrated that during folding, RNA escaped from I_{trap} partitions between M and N, with 10% folds to N and the rest becomes M (*58, 83*). The transitions from I_{trap} and M to N were monitored by activity assay and showed substantial differences between each other. The refolding rate from I_{trap} is 1.5 min^{-1} at 37 °C and 10 mM Mg^{2+}, much faster than that of M (0.002 min^{-1}) as showed in Figure 3.2 (*58, 83*). Moreover, the refolding from I_{trap} is independent on Mg^{2+} concentration, while the escape from M is decreasing largely with elevated Mg^{2+} (5-100 mM) (*83*). Nevertheless, it is unclear what properties of I_{trap} and M lead to these different tracts.

To dissect the physical origin, we first applied denaturant urea in the refolding process. Folding from I_{trap} was monitored by initiating Mg^{2+}-induced folding with trace radio-labeled substrate (50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+} at 37 °C) since I_{trap} accumulates immediately after metal ion addition. Folding from I_{trap} is the rate-limiting step in this process because it is much slower than the substrate cleavage. The formation of product thus reflects the rate of RNA escaping from I_{trap}. Addition of urea (0-2 M) accelerates the overall folding moderately with a *m* value of 0.56 kcal $mol^{-1}M^{-1}$, equivalent to breaking up 5-6 base pairs (Figure A1). This effect is quite small compared to the disruption of 23 base pairs in M (*m* = 1.7 kcal $mol^{-1}M^{-1}$) (*86*), in analogous to the less Mg^{2+} dependence of the former process. Study using oligonucleotide hybridization have detected a folding rate of 1 min^{-1} and urea dependence at 0.45 kcal $mol^{-1}M^{-1}$ under similar conditions (*50, 80*), therefore probably has been followed the same process of I_{trap} transition shown here. Time resolved hydroxyl radical footprinting experiment performed

previously also showed the folding of this step since it gave similar folding rate and urea dependence (*79, 144, 145*).

In the next step, we measure the folding kinetics of mutants that disrupt long-range tertiary interactions by activity assay to investigate the nature of transitions from I_{trap}. Conditions were varied a little compared with those described above to slow folding and show larger effect (50 mM Na-MOPS, pH 7.0, 50 mM Mg^{2+} at 11 °C or 25 °C). The transition from M to N was also performed side by side for better comparison. In most mutants, transition from I_{trap} are only modestly fastened (5-10 fold) (Figure A2), whereas the effect is much stronger in transition from M (results are similar to Ref. (*86*). The disruption of L9.1 and L2.1, though, gave no accelerated refolding rate, indicating that the P13 contact formed between L9.1 and L2.1 remained intact from I_{trap} to N, whereas the other four tertiary contacts may open and close to form the native structure. We also found that despite the impact of mutations on the refolding kinetics, partitioning from I_{trap} to M and N merely changed (within 2 fold) (data not shown), emphasizing the existence of I_{commit} and two separate steps between I_{trap} and N.

The above text contents and following figures were adapted from published paper:

Wan, Y., Suh, H., Russell, R., and Herschlag, D. (2010) Multiple unfolding events during native folding of the Tetrahymena group I ribozyme, *J Mol Biol 400*, 1067-1077.

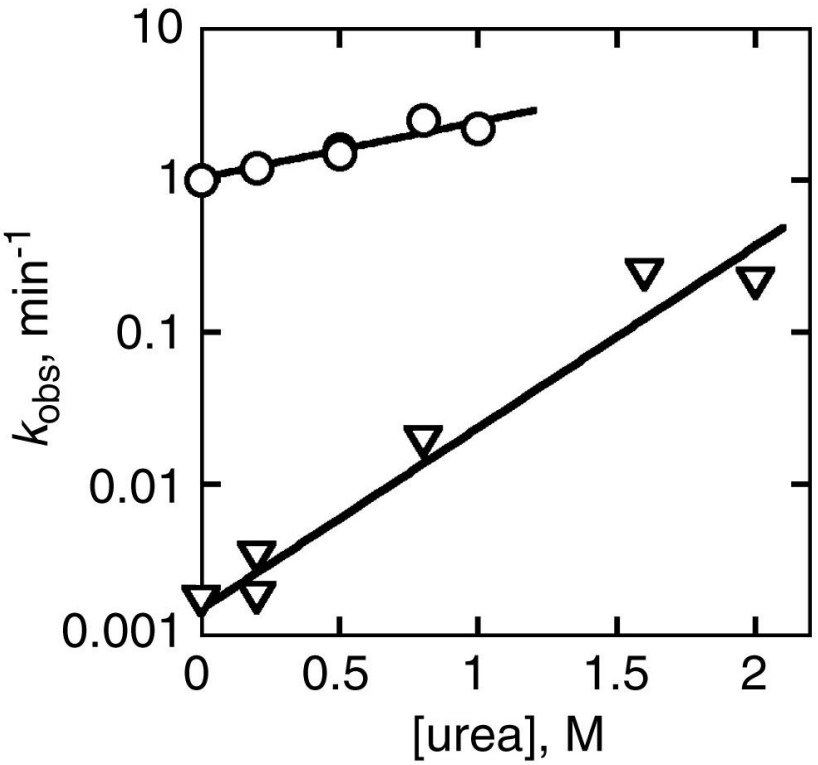

Figure A1: Acceleration by urea of I_{trap} folding to the native state

The data for N (○) is compared with data for re-folding of M under the same conditions at 37 °C and 10 mM Mg^{2+} (▽, data from Ref. (*86*). Across the range of urea concentrations shown for measurements of I_{trap} folding, the rate constant was much larger than that for re-folding of M, allowing a robust determination of the rate constant for the transition from I_{trap} to the native state. At higher urea concentrations (1–2 M), these rate constants become more similar to each other, as shown by the convergence of the lines, such that only the re-folding of M can be measured accurately.

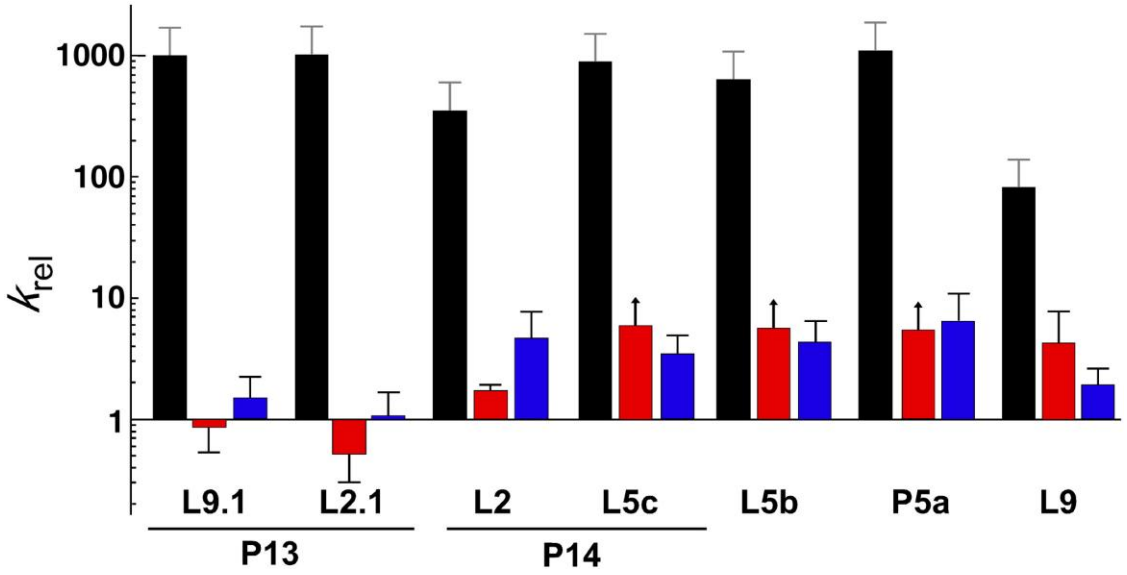

Figure A2: Effects of tertiary contact disruptions on re-folding of I_{trap} and M

Re-folding to the native state from I_{trap} and M. Black bars, re-folding of M to the native state (25 °C, 50 mM Mg^{2+}). Red bars, re-folding of I_{trap} to the native state under the same conditions. Blue bars, re-folding of I_{trap} under conditions that slow folding and therefore allow detection of the larger effects (11 °C, 50 mM Mg^{2+}). Data are shown relative to the corresponding values for the wild-type ribozyme, which were measured in side-by-side reactions. These values are as follows: 2×10^{-5} min^{-1} for re-folding of M, 1.5 ± 0.4 min^{-1} for re-folding of I_{trap} at 25 °C, and 0.30 ± 0.14 min^{-1} for re-folding of I_{trap} at 11 °C. Measurements for folding from I_{trap} were performed at least three times and are shown with error bars corresponding to the standard deviation from these measurements. For some ribozyme variants, only a lower limit for the folding rate constant from I_{trap} was determined at 25 °C, as indicated by upward-pointing arrows above the red bars. For re-folding from M, measurements were performed once, giving relative values that were qualitatively similar to previous measurements performed at 37 °C and 10 mM Mg^{2+} (*86*). The gray error bars show the expected uncertainties based on previous measurements of re-folding, approximately 1.5- to 2-fold.

List of abbreviations

Alt P3– The alternative pairing between J8/7 and the 3'stand of P3 in the *Tetrahymena* group I ribozyme, instead of the native P3
DMS– Dimethyl sulfate
DTT– Dithiothreitol
$E^{\Delta P5abc}$– Variant of the *Tetrahymena* ribozyme which lacks the P5abc peripheral elements
EDTA– Ethylenediaminetetraacetic acid
FRET– Fluorescence resonance energy transfer
G– Guanosine
I_{trap}– The early folding intermediate in the *Tetrahymena* ribozyme folding pathway
L-21– Wild type ribozyme version from self-splicing *Tetrahymena* group I intron that lacks the first 21 nucleotides
M– The long-lived misfolded specie in the *Tetrahymena* ribozyme folding pathway
MOPS– 3-(N-morpholino) propanesulfonic acid
mRNA– Messenger RNA
N– The native structure of the *Tetrahymena* ribozyme
NMR– Nuclear magnetic resonance
P5abc– a peripheral helix extension from the P5 helix in the group I intron. It is composed of three short helices, P5a, P5b and P5c.
PAGE– Polyacrylamide gel electrophoresis
RMSD– Root mean square deviations
RNP– Ribonucleoprotein
rRNA– Ribosomal RNA
tRNA– Transfer RNA
SAXS– Small angle X-ray scattering
SAFA– Semi-Austomated Footprint Analysis
SHAPE– Selective 2'-Hydroxyl Acylation and Primer Extension
TBE buffer– 100 mM Tris, 83 mM boric acid and 1 mM EDTA
TE buffer– 10 mM Tris and 1 mM EDTA
TTE buffer– 70 mM Tris, 20 mM Taurine and 0.4 mM EDTA
WT– Wild type

References:

1. Collins, K. (2006) The biogenesis and regulation of telomerase holoenzymes, *Nat Rev Mol Cell Biol 7*, 484-494.
2. Puglisi, J. D., Blanchard, S. C., and Green, R. (2000) Approaching translation at atomic resolution, *Nat Struct Biol 7*, 855-861.
3. Staley, J. P., and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things, *Cell 92*, 315-326.
4. Buhler, M., Haas, W., Gygi, S. P., and Moazed, D. (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing, *Cell 129*, 707-721.
5. Montange, R. K., and Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function, *Annu Rev Biophys 37*, 117-133.
6. Amaral, P. P., Dinger, M. E., Mercer, T. R., and Mattick, J. S. (2008) The eukaryotic genome as an RNA machine, *Science 319*, 1787-1789.
7. Herschlag, D. (1995) RNA chaperones and the RNA folding problem, *J Biol Chem 270*, 20871-20874.
8. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat Struct Biol 4*, 10-19.
9. Russell, R. (2008) RNA misfolding and the action of chaperones, *Front Biosci 13*, 1-20.
10. Cruz, J. A., and Westhof, E. (2009) The dynamic landscapes of RNA architecture, *Cell 136*, 604-609.
11. Tinoco, I., Jr., and Bustamante, C. (1999) How RNA folds, *J Mol Biol 293*, 271-281.
12. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965) Structure of a Ribonucleic Acid, *Science 147*, 1462-1465.
13. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A. (1973) Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain, *Science 179*, 285-288.
14. Fedor, M. J. (2000) Structure and function of the hairpin ribozyme, *J Mol Biol 297*, 269-291.
15. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme, *Nature 372*, 68-74.
16. Vicens, Q., and Cech, T. R. (2006) Atomic level architecture of group I introns revealed, *Trends Biochem Sci 31*, 41-51.
17. Woodson, S. A. (2005) Structure and assembly of group I introns, *Curr Opin Struct Biol 15*, 324-330.
18. Fedorova, O., and Zingler, N. (2007) Group II introns: structure, folding and splicing mechanism, *Biol Chem 388*, 665-678.

19. Michel, F., Costa, M., and Westhof, E. (2009) The ribozyme core of group II introns: a structure in want of partners, *Trends Biochem Sci 34*, 189-199.
20. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell 35*, 849-857.
21. Evans, D., Marquez, S. M., and Pace, N. R. (2006) RNase P: interface of the RNA and protein worlds, *Trends Biochem Sci 31*, 333-341.
22. Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2006) Structure of ribonuclease P--a universal ribozyme, *Curr Opin Struct Biol 16*, 327-335.
23. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) X-ray crystal structures of 70S ribosome functional complexes, *Science 285*, 2095-2104.
24. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution, *Science 292*, 883-896.
25. Chu, V. B., and Herschlag, D. (2008) Unwinding RNA's secrets: advances in the biology, physics, and modeling of complex RNAs, *Curr Opin Struct Biol 18*, 305-314.
26. Turner, D. H., Sugimoto, N., and Freier, S.M. (1990) Thermodynamics and Kinetics of Base-Pairing and of DNA and RNA Self-Assembly and Helix-Coil Transition, In *Nucleic Acids*, pp 201-227, Springer-Verlag, Berlin.
27. Murray, L. J., Arendall, W. B., 3rd, Richardson, D. C., and Richardson, J. S. (2003) RNA backbone is rotameric, *Proc Natl Acad Sci U S A 100*, 13904-13909.
28. Schneider, B., Moravek, Z., and Berman, H. M. (2004) RNA conformational classes, *Nucleic Acids Res 32*, 1666-1677.
29. Draper, D. E., Grilley, D., and Soto, A. M. (2005) Ions and RNA folding, *Annu Rev Biophys Biomol Struct 34*, 221-243.
30. Kwok, L. W., Shcherbakova, I., Lamb, J. S., Park, H. Y., Andresen, K., Smith, H., Brenowitz, M., and Pollack, L. (2006) Concordant exploration of the kinetics of RNA folding from global and local perspectives, *J Mol Biol 355*, 282-293.
31. Russell, R., Millett, I. S., Tate, M. W., Kwok, L. W., Nakatani, B., Gruner, S. M., Mochrie, S. G., Pande, V., Doniach, S., Herschlag, D., and Pollack, L. (2002) Rapid compaction during RNA folding, *Proc Natl Acad Sci U S A 99*, 4266-4271.
32. Shcherbakova, I., Mitra, S., Laederach, A., and Brenowitz, M. (2008) Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs, *Curr Opin Chem Biol 12*, 655-666.
33. Treiber, D. K., and Williamson, J. R. (1999) Exposing the kinetic traps in RNA folding, *Curr Opin Struct Biol 9*, 339-345.
34. Gartland, W. J., and Sueoka, N. (1966) Two interconvertible forms of tryptophanyl sRNA in E. coli, *Proc Natl Acad Sci U S A 55*, 948-956.
35. Lindahl, T., Adams, A., and Fresco, J. R. (1966) Renaturation of transfer ribonucleic acids through site binding of magnesium, *Proc Natl Acad Sci U S A 55*, 941-948.

36. Esteban, J. A., Banerjee, A. R., and Burke, J. M. (1997) Kinetic mechanism of the hairpin ribozyme. Identification and characterization of two nonexchangeable conformations, *J Biol Chem 272*, 13629-13639.
37. Altman, S., and Guerrier-Takada, C. (1986) M1 RNA, the RNA subunit of Escherichia coli ribonuclease P, can undergo a pH-sensitive conformational change, *Biochemistry 25*, 1205-1208.
38. Pan, J., and Woodson, S. A. (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core, *J Mol Biol 280*, 597-609.
39. Walstrum, S. A., and Uhlenbeck, O. C. (1990) The self-splicing RNA of Tetrahymena is trapped in a less active conformation by gel purification, *Biochemistry 29*, 10573-10576.
40. Herschlag, D., and Cech, T. R. (1990) Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site, *Biochemistry 29*, 10159-10171.
41. Adams, A., Lindahl, T., and Fresco, J. R. (1967) Conformational differences between the biologically active and inactive forms of a transfer ribonucleic acid, *Proc Natl Acad Sci U S A 57*, 1684-1691.
42. Ishida, T., and Sueoka, N. (1967) Rearrangement of the secondary structure of the secondary structure of tryptophan sRNA in Escherichia coli, *Proc Natl Acad Sci U S A 58*, 1080-1087.
43. Cole, P. E., and Crothers, D. M. (1972) Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (Escherichia coli), *Biochemistry 11*, 4368-4374.
44. Uhlenbeck, O. C., Chirikjian, J. G., and Fresco, J. R. (1974) Oligonucleotide binding to the native and denatured conformers of yeast transfer RNA-3 Lea, *J Mol Biol 89*, 495-504.
45. Madore, E., Florentz, C., Giege, R., and Lapointe, J. (1999) Magnesium-dependent alternative foldings of active and inactive Escherichia coli tRNA(Glu) revealed by chemical probing, *Nucleic Acids Res 27*, 3583-3588.
46. Esteban, J. A., Walter, N. G., Kotzorek, G., Heckman, J. E., and Burke, J. M. (1998) Structural basis for heterogeneous kinetics: reengineering the hairpin ribozyme, *Proc Natl Acad Sci U S A 95*, 6091-6096.
47. Chadalavada, D. M., Knudsen, S. M., Nakano, S., and Bevilacqua, P. C. (2000) A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme, *J Mol Biol 301*, 349-367.
48. Chadalavada, D. M., Senchak, S. E., and Bevilacqua, P. C. (2002) The folding pathway of the genomic hepatitis delta virus ribozyme is dominated by slow folding of the pseudoknots, *J Mol Biol 317*, 559-575.
49. Woodson, S. A., and Cech, T. R. (1991) Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA, *Biochemistry 30*, 2042-2050.
50. Zarrinkar, P. P., and Williamson, J. R. (1994) Kinetic intermediates in RNA folding, *Science 265*, 918-924.

51. Rook, M. S., Treiber, D. K., and Williamson, J. R. (1999) An optimal Mg(2+) concentration for kinetic folding of the tetrahymena ribozyme, *Proc Natl Acad Sci U S A 96*, 12471-12476.
52. Zarrinkar, P. P., Wang, J., and Williamson, J. R. (1996) Slow folding kinetics of RNase P RNA, *RNA 2*, 564-573.
53. Pan, T., and Sosnick, T. R. (1997) Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity, *Nat Struct Biol 4*, 931-938.
54. Pan, T., Fang, X., and Sosnick, T. (1999) Pathway modulation, circular permutation and rapid RNA folding under kinetic control, *J Mol Biol 286*, 721-731.
55. Fang, X. W., Pan, T., and Sosnick, T. R. (1999) Mg2+-dependent folding of a large ribozyme without kinetic traps, *Nat Struct Biol 6*, 1091-1095.
56. Fang, X. W., Thiyagarajan, P., Sosnick, T. R., and Pan, T. (2002) The rate-limiting step in the folding of a large ribozyme without kinetic traps, *Proc Natl Acad Sci U S A 99*, 8518-8523.
57. Brehm, S. L., and Cech, T. R. (1983) Fate of an intervening sequence ribonucleic acid: excision and cyclization of the Tetrahymena ribosomal ribonucleic acid intervening sequence in vivo, *Biochemistry 22*, 2390-2397.
58. Russell, R., and Herschlag, D. (1999) New pathways in folding of the Tetrahymena group I RNA enzyme, *J Mol Biol 291*, 1155-1167.
59. Pan, J., Thirumalai, D., and Woodson, S. A. (1997) Folding of RNA involves parallel pathways, *J Mol Biol 273*, 7-13.
60. Pan, T., and Sosnick, T. (2006) RNA folding during transcription, *Annu Rev Biophys Biomol Struct 35*, 161-175.
61. Pan, T., Artsimovitch, I., Fang, X. W., Landick, R., and Sosnick, T. R. (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA, *Proc Natl Acad Sci U S A 96*, 9545-9550.
62. Wong, T., Sosnick, T. R., and Pan, T. (2005) Mechanistic insights on the folding of a large ribozyme during transcription, *Biochemistry 44*, 7535-7542.
63. Karpel, R. L., Miller, N. S., and Fresco, J. R. (1982) Mechanistic studies of ribonucleic acid renaturation by a helix-destabilizing protein, *Biochemistry 21*, 2102-2108.
64. Karpel, R. L., Swistel, D. G., Miller, N. S., Geroch, M. E., Lu, C., and Fresco, J. R. (1975) Acceleration of RNA renaturation by nucleic acid unwinding proteins, *Brookhaven Symp Biol*, 165-174.
65. Cordin, O., Banroques, J., Tanner, N. K., and Linder, P. (2006) The DEAD-box protein family of RNA helicases, *Gene 367*, 17-37.
66. Mahen, E. M., Harger, J. W., Calderon, E. M., and Fedor, M. J. (2005) Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast, *Mol Cell 19*, 27-37.
67. Mahen, E. M., Watson, P. Y., Cottrell, J. W., and Fedor, M. J. (2010) mRNA secondary structures fold sequentially but exchange rapidly in vivo, *PLoS Biol 8*, e1000307.

68. Yadava, R. S., Choi, A. J., Lebruska, L. L., and Fedor, M. J. (2001) Hairpin ribozymes with four-way helical junctions mediate intracellular RNA ligation, *J Mol Biol 309*, 893-902.
69. Haugen, P., Simon, D. M., and Bhattacharya, D. (2005) The natural history of group I introns, *Trends Genet 21*, 111-119.
70. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell 31*, 147-157.
71. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinformatics 3*, 2.
72. Suh, S. O., Jones, K. G., and Blackwell, M. (1999) A Group I intron in the nuclear small subunit rRNA gene of Cryptendoxyla hypophloia, an ascomycetous fungus: evidence for a new major class of Group I introns, *J Mol Evol 48*, 493-500.
73. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *J Mol Biol 216*, 585-610.
74. Zaug, A. J., and Cech, T. R. (1987) Self-splicing RNA and an RNA enzyme in Tetrahymena, *J Protozool 34*, 416-417.
75. Guo, F., Gooding, A. R., and Cech, T. R. (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site, *Mol Cell 16*, 351-362.
76. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme, *Chem Biol 3*, 993-1009.
77. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science 273*, 1678-1685.
78. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998) A preorganized active site in the crystal structure of the Tetrahymena ribozyme, *Science 282*, 259-264.
79. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting, *Science 279*, 1940-1943.
80. Treiber, D. K., Rook, M. S., Zarrinkar, P. P., and Williamson, J. R. (1998) Kinetic intermediates trapped by native interactions in RNA folding, *Science 279*, 1943-1946.
81. Treiber, D. K., and Williamson, J. R. (2001) Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction, *J Mol Biol 305*, 11-21.
82. Pan, J., and Woodson, S. A. (1999) The effect of long-range loop-loop interactions on folding of the Tetrahymena self-splicing RNA, *J Mol Biol 294*, 955-965.

83. Russell, R., and Herschlag, D. (2001) Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway, *J Mol Biol 308*, 839-851.
84. Wan, Y., Mitchell, D 3rd, Russell, R. (2009) Catalytic activity as a probe of native RNA folding, *Methods Enzymol 468*, 195-218.
85. Russell, R., Millett, I. S., Doniach, S., and Herschlag, D. (2000) Small angle X-ray scattering reveals a compact intermediate in RNA folding, *Nat Struct Biol 7*, 367-370.
86. Russell, R., Das, R., Suh, H., Travers, K. J., Laederach, A., Engelhardt, M. A., and Herschlag, D. (2006) The paradoxical behavior of a highly structured misfolded intermediate in RNA folding, *J Mol Biol 363*, 531-544.
87. Adams, P. L., Stahley, M. R., Gill, M. L., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a group I intron splicing intermediate, *RNA 10*, 1867-1887.
88. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons, *Nature 430*, 45-50.
89. Golden, B. L., Kim, H., and Chase, E. (2005) Crystal structure of a phage Twort group I ribozyme-product complex, *Nat Struct Mol Biol 12*, 82-89.
90. Vicens, Q., Paukstelis, P. J., Westhof, E., Lambowitz, A. M., and Cech, T. R. (2008) Toward predicting self-splicing and protein-facilitated splicing of group I introns, *RNA 14*, 2013-2029.
91. Tanner, M., and Cech, T. (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium Azoarcus, *RNA 2*, 74-83.
92. Caprara, M. G., Lehnert, V., Lambowitz, A. M., and Westhof, E. (1996) A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core, *Cell 87*, 1135-1145.
93. Paukstelis, P. J., Coon, R., Madabusi, L., Nowakowski, J., Monzingo, A., Robertus, J., and Lambowitz, A. M. (2005) A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface, *Mol Cell 17*, 417-428.
94. Longo, A., Leonard, C. W., Bassi, G. S., Berndt, D., Krahn, J. M., Hall, T. M., and Weeks, K. M. (2005) Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase, *Nat Struct Mol Biol 12*, 779-787.
95. Weeks, K. M., and Cech, T. R. (1996) Assembly of a ribonucleoprotein catalyst by tertiary structure capture, *Science 271*, 345-348.
96. Pichler, A., and Schroeder, R. (2002) Folding problems of the 5' splice site containing the P1 stem of the group I thymidylate synthase intron: substrate binding inhibition in vitro and mis-splicing in vivo, *J Biol Chem 277*, 17987-17993.
97. Zhang, L., Xiao, M., Lu, C., and Zhang, Y. (2005) Fast formation of the P3-P7 pseudoknot: a strategy for efficient folding of the catalytically active ribozyme, *RNA 11*, 59-69.
98. Jiang, Y. F., Xiao, M., Yin, P., and Zhang, Y. (2006) Monovalent cations use multiple mechanisms to resolve ribozyme misfolding, *RNA 12*, 561-566.

99. Zhang, L., Bao, P., Leibowitz, M. J., and Zhang, Y. (2009) Slow formation of a pseudoknot structure is rate limiting in the productive co-transcriptional folding of the self-splicing Candida intron, *RNA 15*, 1986-1992.
100. Chauhan, S., Caliskan, G., Briber, R. M., Perez-Salas, U., Rangan, P., Thirumalai, D., and Woodson, S. A. (2005) RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme, *J Mol Biol 353*, 1199-1209.
101. Rangan, P., Masquida, B., Westhof, E., and Woodson, S. A. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme, *Proc Natl Acad Sci U S A 100*, 1574-1579.
102. Bhattacarya, D., Cannone, J. J., and Gutell, R. R. (2001) Group I intron lateral transfer between red and brown algal ribosomal RNA, *Curr Genet 40*, 82-90.
103. Muller, K. M., Cannone, J. J., Gutell, R. R., and Sheath, R. G. (2001) A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta), *Mol Biol Evol 18*, 1654-1667.
104. Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) Tertiary Motifs in RNA Structure and Folding, *Angew Chem Int Ed Engl 38*, 2326-2343.
105. Reymond, C., Beaudoin, J. D., and Perreault, J. P. (2009) Modulating RNA structure and catalysis: lessons from small cleaving ribozymes, *Cell Mol Life Sci 66*, 3937-3950.
106. Toor, N., Keating, K. S., Taylor, S. D., and Pyle, A. M. (2008) Crystal structure of a self-spliced group II intron, *Science 320*, 77-82.
107. Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D., and Pace, N. R. (2005) Crystal structure of a bacterial ribonuclease P RNA, *Proc Natl Acad Sci U S A 102*, 13392-13397.
108. Zaug, A. J., Been, M. D., and Cech, T. R. (1986) The Tetrahymena ribozyme acts like an RNA restriction endonuclease, *Nature 324*, 429-433.
109. Sigler, P. B. (1975) An analysis of the structure of tRNA, *Annu Rev Biophys Bioeng 4*, 477-527.
110. Fang, X. W., Golden, B. L., Littrell, K., Shelton, V., Thiyagarajan, P., Pan, T., and Sosnick, T. R. (2001) The thermodynamic origin of the stability of a thermophilic ribozyme, *Proc Natl Acad Sci U S A 98*, 4355-4360.
111. Richardson, J. S., and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation, *Proc Natl Acad Sci U S A 99*, 2754-2759.
112. Berezovsky, I. N., Zeldovich, K. B., and Shakhnovich, E. I. (2007) Positive and negative design in stability and thermal adaptation of natural proteins, *PLoS Comput Biol 3*, e52.
113. Havranek, J. J., and Harbury, P. B. (2003) Automated design of specificity in molecular recognition, *Nat Struct Biol 10*, 45-52.
114. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004) Computational redesign of protein-protein interaction specificity, *Nat Struct Mol Biol 11*, 371-379.
115. Bolon, D. N., Grant, R. A., Baker, T. A., and Sauer, R. T. (2005) Specificity versus stability in computational protein design, *Proc Natl Acad Sci U S A 102*, 12724-12729.

116. Guo, F., and Cech, T. R. (2002) Evolution of Tetrahymena ribozyme mutants with increased structural stability, *Nat Struct Biol 9*, 855-861.
117. Guo, F., Gooding, A. R., and Cech, T. R. (2006) Comparison of crystal structure interactions and thermodynamics for stabilizing mutations in the Tetrahymena ribozyme, *RNA 12*, 387-395.
118. Fang, X. W., Srividya, N., Golden, B. L., Sosnick, T. R., and Pan, T. (2003) Stepwise conversion of a mesophilic to a thermophilic ribozyme, *J Mol Biol 330*, 177-183.
119. Johnson, T. H., Tijerina, P., Chadee, A. B., Herschlag, D., and Russell, R. (2005) Structural specificity conferred by a group I RNA peripheral element, *Proc Natl Acad Sci U S A 102*, 10176-10181.
120. Russell, R., and Herschlag, D. (1999) Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme, *RNA 5*, 158-166.
121. Zaug, A. J., Grosshans, C. A., and Cech, T. R. (1988) Sequence-specific endoribonuclease activity of the Tetrahymena ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes, *Biochemistry 27*, 8924-8931.
122. Flores, S. C., Wan, Y., Russell, R., and Altman, R. B. (2010) Predicting RNA structure by multiple template homology modeling, *Pac Symp Biocomput*, 216-227.
123. Russell, R., Tijerina, P., Chadee, A. B., and Bhaskaran, H. (2007) Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme, *Biochemistry 46*, 4951-4961.
124. Doherty, E. A., and Doudna, J. A. (1997) The P4-P6 domain directs higher order folding of the Tetrahymena ribozyme core, *Biochemistry 36*, 3159-3169.
125. Engelhardt, M. A., Doherty, E. A., Knitt, D. S., Doudna, J. A., and Herschlag, D. (2000) The P5abc peripheral element facilitates preorganization of the tetrahymena group I ribozyme for catalysis, *Biochemistry 39*, 2639-2651.
126. Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) Mapping adenines, guanines, and pyrimidines in RNA, *Nucleic Acids Res 4*, 2527-2538.
127. Huang, Z., and Szostak, J. W. (1996) A simple method for 3'-labeling of RNA, *Nucleic Acids Res 24*, 4360-4361.
128. van der Horst, G., Christian, A., and Inoue, T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA, *Proc Natl Acad Sci U S A 88*, 184-188.
129. Doherty, E. A., Herschlag, D., and Doudna, J. A. (1999) Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme, *Biochemistry 38*, 2982-2990.
130. Banerjee, A. R., Jaeger, J. A., and Turner, D. H. (1993) Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure, *Biochemistry 32*, 153-163.
131. Russell, R., Zhuang, X., Babcock, H. P., Millett, I. S., Doniach, S., Chu, S., and Herschlag, D. (2002) Exploring the folding landscape of a structured RNA, *Proc Natl Acad Sci U S A 99*, 155-160.

132. Sattin, B. D., Zhao, W., Travers, K., Chu, S., and Herschlag, D. (2008) Direct measurement of tertiary contact cooperativity in RNA folding, *J Am Chem Soc 130*, 6085-6087.
133. Wan, Y., Suh, H., Russell, R., and Herschlag, D. (2010) Multiple unfolding events during native folding of the Tetrahymena group I ribozyme, *J Mol Biol 400*, 1067-1077.
134. Silverman, S. K., and Cech, T. R. (1999) Energetics and cooperativity of tertiary hydrogen bonds in RNA structure, *Biochemistry 38*, 8691-8702.
135. Blose, J. M., Silverman, S. K., and Bevilacqua, P. C. (2007) A simple molecular model for thermophilic adaptation of functional nucleic acids, *Biochemistry 46*, 4232-4240.
136. Bhaskaran, H., and Russell, R. (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone, *Nature 449*, 1014-1018.
137. Sclavi, B., Woodson, S., Sullivan, M., Chance, M. R., and Brenowitz, M. (1997) Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding, *J Mol Biol 266*, 144-159.
138. Shcherbakova, I., Mitra, S., Beer, R. H., and Brenowitz, M. (2006) Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins, *Nucleic Acids Res 34*, e48.
139. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D., and Altman, R. B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments, *RNA 11*, 344-354.
140. Inoue, T., and Cech, T. R. (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase, *Proc Natl Acad Sci U S A 82*, 648-652.
141. Tijerina, P., Mohr, S., and Russell, R. (2007) DMS footprinting of structured RNAs and RNA-protein complexes, *Nat Protoc 2*, 2608-2623.
142. Qu, H. L., Michot, B., and Bachellerie, J. P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA, *Nucleic Acids Res 11*, 5903-5920.
143. Moazed, D., Stern, S., and Noller, H. F. (1986) Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension, *J Mol Biol 187*, 399-416.
144. Laederach, A., Shcherbakova, I., Liang, M. P., Brenowitz, M., and Altman, R. B. (2006) Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule, *J Mol Biol 358*, 1179-1190.
145. Shcherbakova, I., and Brenowitz, M. (2005) Perturbation of the hierarchical folding of a large RNA by the destabilization of its Scaffold's tertiary structure, *J Mol Biol 354*, 483-496.
146. Murphy, F. L., and Cech, T. R. (1993) An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme, *Biochemistry 32*, 5291-5300.

147. Ditzler, M. A., Rueda, D., Mo, J., Hakansson, K., and Walter, N. G. (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation, *Nucleic Acids Res 36*, 7088-7099.
148. Cech, T. R., and Bass, B. L. (1986) Biological catalysis by RNA, *Annu Rev Biochem 55*, 599-629.
149. Keenan, R. J., Freymann, D. M., Stroud, R. M., and Walter, P. (2001) The signal recognition particle, *Annu Rev Biochem 70*, 755-775.
150. Holmes, K. L., and Culver, G. M. (2005) Analysis of conformational changes in 16 S rRNA during the course of 30 S subunit assembly, *J Mol Biol 354*, 340-357.
151. Orr, J. W., Hagerman, P. J., and Williamson, J. R. (1998) Protein and Mg(2+)-induced conformational changes in the S15 binding site of 16 S ribosomal RNA, *J Mol Biol 275*, 453-464.
152. Duncan, C. D., and Weeks, K. M. (2008) SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA, *Biochemistry 47*, 8504-8513.
153. Treiber, D. K., and Williamson, J. R. (2001) Beyond kinetic traps in RNA folding, *Curr Opin Struct Biol 11*, 309-314.
154. Flores, S. C., and Altman, R. B. (2010) Turning limited experimental information into 3D models of RNA, *RNA 16*, 1769-1778.
155. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.
156. Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature, *J Comput Chem 23*, 600-609.
157. Svergun, D., Barberato, C. and Koch, M. H. (1995) CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates., *J. Appl. Crystallogr 28*, 768-773.
158. Serra, M. J., and Turner, D. H. (1995) Predicting thermodynamic properties of RNA, *Methods Enzymol 259*, 242-261.
159. Narlikar, G. J., Khosla, M., Usman, N., and Herschlag, D. (1997) Quantitating tertiary binding energies of 2' OH groups on the P1 duplex of the Tetrahymena ribozyme: intrinsic binding energy in an RNA enzyme, *Biochemistry 36*, 2465-2477.
160. Costa, M., and Michel, F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution, *Embo J 16*, 3289-3302.
161. Gutell, R. R., Lee, J. C., and Cannone, J. J. (2002) The accuracy of ribosomal RNA comparative structure models, *Curr Opin Struct Biol 12*, 301-310.
162. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix, *Nature 372*, 111-113.
163. Costa, M., and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs, *EMBO J 14*, 1276-1285.
164. Flores, S. C., and Altman, R. B. (2010) Turning limited experimental information into 3D models of RNA, *RNA*.

165. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Res 30*, 3497-3531.
166. Lipfert, J., and Doniach, S. (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes, *Annu Rev Biophys Biomol Struct 36*, 307-327.

Vita

Yaqi Wan was born in Wuhan, China. She attended Wuhan University in 2000 and obtained her B.S degree in biotechnology after four years. She then went to the institute of cellular and molecular biology at the University of Texas at Austin to pursue her doctoral degree in Dr. Rick Russell's lab.

Permanent address: c/o Rick Russell, MBB2.212, A2400 University of Texas, Austin, TX 78712

This dissertation was typed by Yaqi Wan.